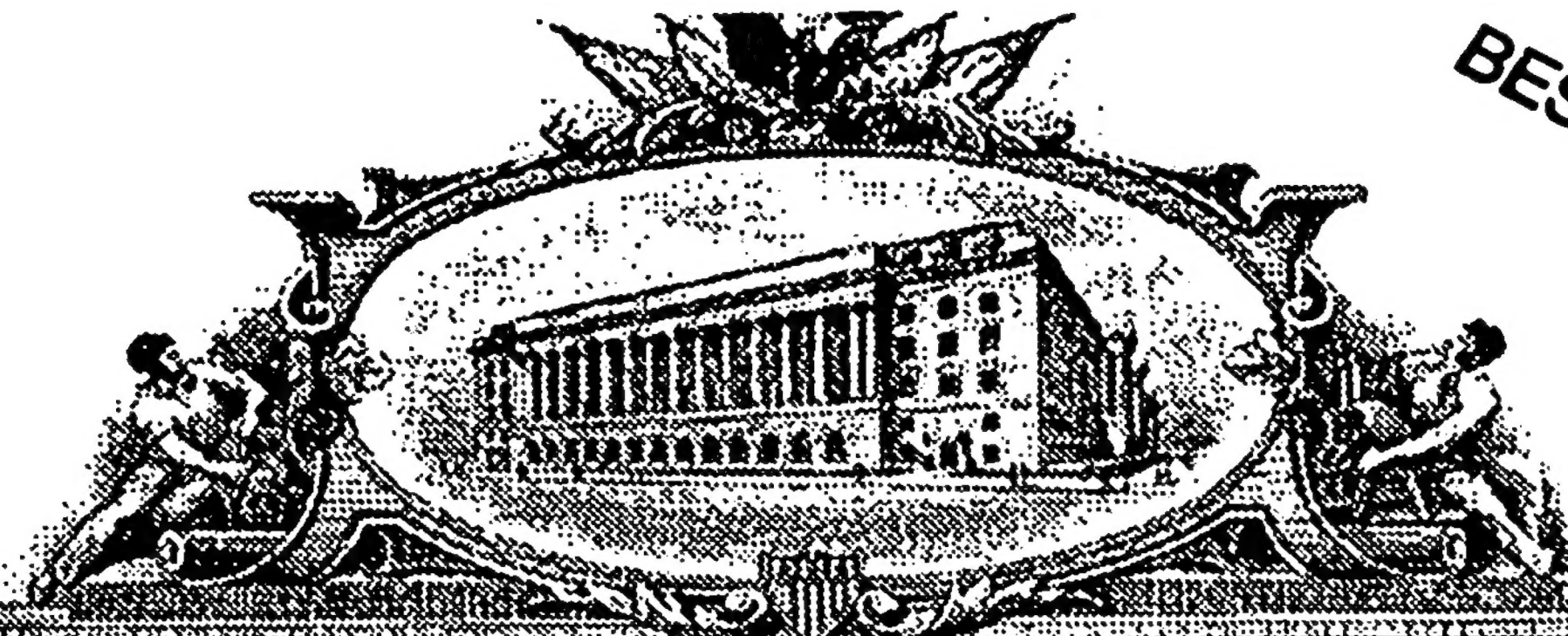


BEST AVAILABLE COPY



1268994

# THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

*January 04, 2005*

**THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.**

**APPLICATION NUMBER: 60/491,822**

**FILING DATE: July 31, 2003**

**RELATED PCT APPLICATION NUMBER: PCT/US04/24868**



Certified By

Jon W Dudas

Under Secretary  
of Commerce for Intellectual Property  
and Acting Director of the  
United States Patent and Trademark Office



17364 U.S. PTO

"Express Mail" Mailing Label No. EV 351 355 394 US

Date of Deposit: July 31, 2003

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR §1.10 on the date indicated above and is addressed to the Commissioner of Patents & Trademarks, Washington, D.C. 20231.

*Rebecca M. Hale*  
Rebecca M. Hale

Date: July 31, 2003

15535 U.S. PT  
60/491822  
07/31/03

**PROVISIONAL APPLICATION COVER SHEET**

This is a request for a PROVISIONAL APPLICATION under 37 CFR 1.53(c).

Customer No. 27476		Docket No. 20663.001	Type a plus sign (+) inside this box → +
INVENTOR(S)/APPLICANT(S)			
LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Grandl	Guido		
Telford	John		
Bensi	Giuliano		
TITLE OF INVENTION (280 characters max)			
Immunogenic Compositions for <i>Streptococcus pyogenes</i>			
CORRESPONDENCE ADDRESS			
Rebecca M. Hale CHIRON CORPORATION Intellectual Property - R440 P.O. Box 8097 Emeryville			
STATE: California	ZIP CODE: 94662-8097	COUNTRY: USA	
ENCLOSED APPLICATION PARTS (check all that apply)			
<input checked="" type="checkbox"/> Specification		Number of Pages: 58	
<input type="checkbox"/> Drawing(s)		Number of Pages:	
<input type="checkbox"/> Small Entity Statement			
<input type="checkbox"/> Other (specify)			
METHOD OF PAYMENT (check one)			
<input checked="" type="checkbox"/> A check or money order is enclosed to cover the Provisional filing fees			
<input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge any additional fees and credit Deposit Account Number 03-1664.		PROVISIONAL FILING FEE AMOUNT ENCLOSED \$160.00 CHECK NO. 8182	

The invention was made by an agency of the United States Government or under a contract with an agency of the United States government.

☒ No

☐ Yes, the name of the U.S. Government agency and the Government contract number are:

July 31, 2003  
CHIRON CORPORATION  
Intellectual Property - R440  
P.O. Box 8097  
Emeryville, CA 94662-8097  
(510) 923-3179 - (510) 655-3542 (fax)

Respectfully submitted,  
By: *Rebecca M. Hale*  
Rebecca M. Hale  
Attorney for Applicants  
Reg. No. 45,680

## IMMUNOGENIC COMPOSITIONS FOR *STREPTOCOCCUS PYOGENES*

All documents cited herein are incorporated by reference in their entirety.

### TECHNICAL FIELD

This invention is in the fields of immunology and vaccinology. In particular, it relates to antigens  
5 derived from *Streptococcus pyogenes* and their use in immunisation.

### BACKGROUND ART

Group A streptococcus ("GAS", *S.pyogenes*) is a frequent human pathogen, estimated to be present in between 5-15% of normal individuals without signs of disease. When host defences are compromised, or when the organism is able to exert its virulence, or when it is introduced to  
10 vulnerable tissues or hosts, however, an acute infection occurs. Related diseases include puerperal fever, scarlet fever, erysipelas, pharyngitis, impetigo, necrotising fasciitis, myositis and streptococcal toxic shock syndrome.

Although *S.pyogenes* may be treated using antibiotics, a prophylactic vaccine to prevent the onset of disease is desired. Efforts to develop such a vaccine have been ongoing for many decades.

15 While various GAS vaccine approaches have been suggested and some approaches are currently in clinical trials, to date, there are no GAS vaccines available to the public.

It is an object of the invention to provide further and improved compositions for providing immunity against GAS disease and/or infection. The compositions are based on a combination of two or more (e.g. three or more) GAS antigens.

### 20 DISCLOSURE OF THE INVENTION

Applicants have discovered a group of thirty GAS antigens that are particularly suitable for immunisation purposes, particularly when used in combinations. The invention therefore provides an immunogenic composition comprising a combination of GAS antigens, said combination consisting of two to thirty-one GAS antigens of a first antigen group, said first antigen group consisting of: GAS  
25 117, GAS 130, GAS 277, GAS 236, GAS 40, GAS 389, GAS 504, GAS 509, GAS 366, GAS 159, GAS 217, GAS 309, GAS 372, GAS 039, GAS 042, GAS 058, GAS 290, GAS 511, GAS 533, GAS 527, GAS 294, GAS 253, GAS 529, GAS 045, GAS 095, GAS 193, GAS 137, GAS 084, GAS 384, GAS 202, and GAS 057. These antigens are referred to herein as the 'first antigen group'.

Preferably, the combination of GAS antigens consists of three, four, five, six, seven, eight, nine, or ten  
30 GAS antigens selected from the first antigen group. Preferably, the combination of GAS antigens consists of three, four, or five GAS antigens selected from the first antigen group.

GAS 40 and GAS 117 are particularly preferred GAS antigens. Preferably, the combination of GAS antigens includes either or both of GAS 40 and GAS 117. Representative examples of some of these antigen combinations are discussed below.



The combination of GAS antigens may consist of three GAS antigens selected from the first antigen group. Accordingly, in one embodiment, the combination of GAS antigens consists of GAS 40, GAS 117 and a third GAS antigen selected from the first antigen group. In another embodiment, the combination of GAS antigens consists of GAS 40 and two additional GAS antigens selected from the first antigen group. In another embodiment, the combination of GAS antigens consists of GAS 117 and two additional GAS antigens selected from the first antigen group.

The combination of GAS antigens may consist of four GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40, GAS 117 and two additional GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40 and three additional GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 117 and three additional antigens selected from the first antigen group.

The combination of GAS antigens may consist of five GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40, GAS 117 and three additional GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40 and four additional GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 117 and four additional GAS antigens selected from the first antigen group.

The combination of GAS antigens may consist of eight GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40, GAS 117 and six additional GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40 and seven additional GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 117 and seven additional GAS antigens selected from the first antigen group.

The combination of GAS antigens may consist of ten GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40, GAS 117 and eight additional GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 40 and nine additional GAS antigens selected from the first antigen group. In one embodiment, the combination of GAS antigens consists of GAS 117 and nine additional GAS antigens selected from the first antigen group.

Each of the GAS antigens of the first antigen group are described in more detail below. Genomic sequences of at least three GAS strains are publicly available. The genomic sequence of an M1 GAS strain is reported at Ref. 1. The genomic sequence of an M3 GAS strain is reported at Ref. 2. The genomic sequence of an M18 GAS strain is reported at Ref. 3. Preferably, the GAS antigens of the invention comprise polynucleotide or amino acid sequence of an M1, M3 or M18 GAS strains. More preferably, the GAS antigens of the invention comprise a polynucleotide or amino acid sequence of an M1 strain.



**(1) GAS 117**

GAS 117 corresponds to M1 GenBank accession numbers GI:13621679 and GI:15674571, to M3 GenBank accession number GI:21909852, to M18 GenBank accession number GI: 19745578, and is also referred to as 'Spy0448' (M1), 'SpyM3\_0316' (M3), and 'SpyM18\_0491' (M18). Examples of amino acid and polynucleotide sequences of GAS 117 of an M1 strain are set forth below:

**SEQ ID NO: 1**

MTLKKHYLLSLLALVTGAAFNTSQS~~VSAQVYS~~NEGYHQHLTDEKSHLQYSKD~~NAQLQLRN~~ILDGYQND  
LGRHYSSYYYNLRTVMGLSSEQDIEKH~~YSELK~~NKLHDMYNHY

**SEQ ID NO: 2**

ATGACACTAAAAAACA~~CTATTATCTTCTCAGCCTGCTAGCTCTTGTAACGGTTGGTGCTGCCTTTAACA~~  
CAAGCCAGAGTGTCA~~GTGCACAAGTTTATAGCAATGAAGGGTATCACCAGCATTGACTGATGAAAAATC~~  
ACACCTGCAATATAGTAAAGACAACGCACA~~ACTTCAATTGAGAAATATCCTTGACGGCTACCAAAATGAC~~  
CTAGGGAGACACTACTCTAGCTATTATTACTACAACCTAAGAACCGTTATGGGACTATCAAGTGAGCAAG  
ACATTGAAAAACACTATGAAGAGCTTAAGAACAAGTTACATGATATGTACAATCATTATTAA

Preferred GAS 117 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 1; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 1, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 or more). These GAS 117 proteins include variants (e.g. allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 1. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 1. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 1. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 1 is removed. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

**(2) GAS 130**

GAS 130 corresponds to M1 GenBank accession numbers GI:13621794 and GI:15674677, to M3 GenBank accession number GI: 21909954, to M18 GenBank accession number GI: 19745704, and is also referred to as 'Spy0591' (M1), 'SpyM3\_0418' (M3), and 'SpyM18\_0660' (M18). GAS 130 has potentially been identified as a putative protease. Examples of amino acid and polynucleotide sequences of GAS 130 of an M1 strain are set forth below:

**SEQ ID NO: 3**

MSHMKRPEVLS~~PAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGIDYAHARGAKVYVAA~~  
NMVTHEGNEIGAGEWFRQLRDMGLDAVIVSDPALIVICSTEAPGLEIHLSTQASSTNYETFEFWKAMGLT  
RVVLAREVNMAELAEIRKRTDVEIEAFVHGAMCISYSGRCVLSNHMSHRDANRGGCSQSCRWKYDLYDMP  
FGGERRSLKGEIPEDYSMSVDMCMIDHIPDLIENGVD~~SLKIEGRMKS~~IHYVSTVTN~~CYKAAVGAYMES~~P  
EAFYAIKEELIDELWKVAQRELATGFYYGIPTENEQLFGARRKIPOYKFVGEVVA~~FDSASMTATIRQRNV~~  
IMEGDRIECYGPGRHFETVVKDLHDADGQKIDRAPNPMELLTISLPREVKPGDMIRACKEGLVNLYQKD  
GTSKTVRT

**SEQ ID NO: 4**

ATGTCACATATGAAAAACGTCCCGAGGTCTTATCACCTGCTGGAACACTTGAAAAATTAAAGTTGCGA  
TTGACTATGGCGCAGATGCTGTTTTTGTGGAGGGCAGGCCTATGGCCTAAGAAGCCGCGCTGGTAACTT



CTCTATGGAAGAATTGCAAGAAGGCATTGATTATGCACATGCGCGTGGAGCTAAGGTCTATGTTGCTGCT  
AACATGGTTACCCACGAAGGGAACGAAATTGGTGCAGGCGAGTGGTTTCGTCAACTGCGTGATATGGGGC  
TTGATGCGGTCATTGTTTCAGATCCAGCCTTGATTGTTATTTGTTCAACAGAAGCCCCAGGTTTGAAAT  
5 TCATTTGTCAACGCAAGCTTCATCTACCAATTACGAGACCTTTGAATTTTGGAAAGCCATGGGCTTGACC  
CGAGTTGTTTTAGCTCGCGAGGTTAATATGGCCGAGTTAGCAGAAATCCGCAAGCGGACAGATGTGGAAA  
TTGAAGCCTTTGTCCATGGAGCCATGTGTATCTCTTATTCAGGCCGCTGTGTTTTGTCAAACCACATGAG  
10 TCACCGTGATGCCAACAGGGGCGGCTGCTCACAGTCTTGCCGCTGGAAGTATGATTTGTATGACATGCCA  
TTTGGAGGAGAGCGCCGCTCCTTAAAGGGGAAATTCAGAAGACTATTCTATGTCCTCTGTTGACATGT  
GTATGATTGACCATATTCCTGACCTGATTGAAAATGGGGTTGATAGCTTAAAAATTGAAGGCCGAATGAA  
ATCTATCCACTACGTCTCAACCGTAACCACTGTTACAAGGCGGCTGTAGGTGCTTACATGGAAAGCCCA  
GAAGCTTTTTATGCTATCAAAGAGGAATTGATTGACGAGTTGTGGAAAGGTTGCCAGCGCGAGTTGGCTA  
CAGGTTTTTACTATGGTATCCCAACTGAAAATGAACAATTATTTGGTGCTCGCCGCAAAATTCCACAATA  
TAAATTTGTCGGAGAAGTAGTTGCCTTTGACTCAGCTAGCATGACAGCGACCATTCGTCAGCGTAATGTC  
ATCATGGAAGGCGATCGGATTGAATGTTATGGACCAGGTTTCCGTCATTTTGAAACGGTTGTTAAGGACT  
15 TACATGATGCGGATGGCCAAAAGATTGACCGTGCCCCAAATCCAATGGAAGTCTTAACCATCTCTTTACC  
GAGAGAAGTTAAGCCAGGGGATATGATTAGGGCTTGCAAGGAAGGTCTGGTTAACCTCTATCAAAAAGAT  
GGCACCAGTAAACTGTTAGAACATAG

Preferred GAS 130 proteins for use with the invention comprise an amino acid sequence: (a) having  
20 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 3; and/or (b) which is a fragment of at least  $n$   
consecutive amino acids of SEQ ID NO: 3, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
30, 35, 40, 50, 60, 70, 80, 90, 100, 150, or more). These GAS 130 proteins include variants (e.g.  
allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 3. Preferred fragments  
25 of (b) comprise an epitope from SEQ ID NO: 3. Other preferred fragments lack one or more amino  
acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more  
amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID  
NO: 3. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of  
a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

### 30 (3) GAS 277

GAS 277 corresponds to M1 GenBank accession numbers GI:13622962 and GI:15675742, to M3  
GenBank accession number GI: 21911206, to M18 GenBank accession number GI: 19746852, and is  
also referred to as 'Spy1939' (M1), 'SpyM3\_1670' (M3), and 'SpyM18\_2006' (M18). Amino acid  
and polynucleotide sequences of GAS 277 of an M1 strain are set forth below:

### 35 SEQ ID NO: 5

MTTMQKTIISLLSLALLIGLLGTSGKAISVYAQDQHTDNVIAESTISQVSVEASMRGTEPYIDATVTTDQP  
VRQPTQATITLKDASDNTINSWVYTMAAQRRFTAWFDLTGQKSGDYHVTVTVHTQEKAVTGQSGTVHFD  
QNKARKTPTNMQQKDTSKAMTNSVDVDTKAQTNQSANQEIDSTSNPFRSATNHRSTSLKRSTKNEKLTP  
40 ASNSQKNGSNKTKMLVDKEEVKPTSKRGFPWVLLGLVLSLAAGLFIAIQKVSRRK

### 40 SEQ ID NO: 6

ATGACAACTATGCAAAAAACAATTAGCTTATTATCACTAGCTTTACTTATTGGTTTGCTGGGGACTTCTG  
GCAAAGCCATATCTGTGTATGCACAAGATCAGCACACTGATAATGTTATAGCTGAATCAACTATTAGTCA  
GGTCAGTGTTGAAGCCAGTATGCGTGGAACAGAACCTTATATTGATGCTACAGTCACCACAGATCAACCT  
45 GTCAGACAACCAACTCAGGCAACGATAAAGACTTAAAGACGCTAGTGATAATACTATTAATAGTTGGGTAT  
ATACTATGGCAGCGCAACAGCGTCGTTTTACAGCTTGGTTGATTAACTGGACAAAAGAGTGGTGACTA  
TCATGTAAGTGTACCGTTCACTACTCAAGAAAAGGCAGTAAGTGGTCAATCAGGAAGTGTTCATTTTGAT  
CAAAACAAAGCTAGAAAAACACCAACTAATATGCAACAAAAGGATACTTCTAAAGCAATGACGAATTCAG  
TCGATGTAGACACAAAAGCTCAAACAAATCAATCAGCTAACCAAGAAATAGATTCTACTTCAAATCCTTT  
50 CAGATCAGCTACTAATCATCGATCAACTTCCTTAAAGCGATCTACTAAAAATGAGAACTTACACCAACT  
GCTAGTAATAGCCAAAAAACGGTAGCAACAAGACAAAAATGCTAGTGACAAAGAGGAAGTAAACCTA



CTTCAAAAAGAGGATTCCCTTGGGTCTTATTAGGTCTAGTAGTCAGTTTAGCTGCAGGTTTATTTATAGC  
TATTCAAAAAGTATCTAGACGAAAATAA

Preferred GAS 277 proteins for use with the invention comprise an amino acid sequence: (a) having  
5 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 5; and/or (b) which is a fragment of at least  $n$   
consecutive amino acids of SEQ ID NO: 5, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
30, 35, 40, 50, 60, 70, 80, 90, 100, or more). These GAS 277 proteins include variants (e.g. allelic  
variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 5. Preferred fragments of (b)  
10 comprise an epitope from SEQ ID NO: 5. Other preferred fragments lack one or more amino acids  
(e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino  
acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 5. For  
example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 5  
is removed. Other fragments omit one or more domains of the protein (e.g. omission of a signal  
15 peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

#### (4) GAS 236

GAS 236 corresponds to M1 GenBank accession numbers GI:13622264 and GI:15675106, M3  
GenBank accession number GI: 21910321, and to M18 GenBank accession number GI: 19746075,  
and is also referred to as 'Spy1126' (M1), 'SpyM3\_0785' (M3), and 'SpyM18\_1087' (M18). Amino  
20 acid and polynucleotide sequences of GAS 236 from an M1 strain are set forth below:

#### SEQ ID NO: 7

MTQMNYTGKVKRVAIIANGKYQSKRVASKLFSVFKDDPDFYLSKKNPDIVISIGGDGMLLSAFHMYEKEL  
DKVRFVGIHTGHLGFYTDYRDFEVDKLDNLRKDKGEQISYPILKVAITLDDGRVVKARALNEATVKRIE  
KTMVADVIIINHVKFESFRGDGISVSTPTGSTAYNKS LGGAVLHPTIEALQLTEISSLNRRVFRTLGSII  
25 I PKKDKIELVPKRLGIYTI SIDNKTYQLKNVTKVEYFIDDEKIHVSSPSHTSPWERVKDAFIGEIDS

#### SEQ ID NO: 8

ATGACACAGATGAATTATACAGGTAAGGTAAAACGAGTTGCTATTATTGCAAATGGTAAGTACCAAAGTA  
AACGCGTCGCTCCAACTTTTCTCCGTATTTAAAGATGATCCTGATTTCTATCTTTCAAAGAAAAATCC  
30 GGATATTGTGATTTCTATTGGCGGAGATGGGATGCTCTTATCTGCCTTTCACATGTATGAAAAAGAATTA  
GATAAGGTACGTTTTGTAGGAATCCACACCGGTCATCTTGGCTTTTATACCGATTATAGGGATTTTGAAG  
TTGATAAATTAATTGATAATTTAAGAAAAGACAAGGGAGAACAATCTCTTATCCGATTTTAAAAGTTGC  
TATTACTTTAGATGATGGTCGTGTGGTTAAAGCGCGTGCTTTGAATGAAGCGACGGTTAAGCGTATTGAA  
AAAACGATGGTAGCAGATGTTATTATTAACCATGTCAAATTTGAAAGCTTCCGAGGTGATGGGATTTTCAG  
35 TATCGACCCCGACAGGGAGCACAGCCTACAATAAATCTTTAGGTGGTGCTGTCTTGCATCCGACGATTGA  
AGCGCTGCAATTGACGGAAATTTCCAGTCTTAATAACCGTGCTTTAGAACCTTGGGCTCATCAATCATT  
ATTCCCAAAAAGATAAGATTGAGTTAGTGCCAAAACGATTAGGAATTTATACCATTTCATTGATAATA  
AAACCTATCAGTTAAAAAATGTGACGAAGGTGGAGTATTTTATCGACGATGAGAAAATTCATTTTGTTTC  
CTCTCCGAGTCATACGAGCTTTTGGGAAAGGGTCAAGGATGCCTTTATTGGAGAGATTGACTCATGA  
40

Preferred GAS 236 proteins for use with the invention comprise an amino acid sequence: (a) having  
50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 7; and/or (b) which is a fragment of at least  $n$   
consecutive amino acids of SEQ ID NO: 7, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
30, 35, 40, 50, 60, 70, 80, 90, 100, 150 or more). These GAS 236 proteins include variants (e.g. allelic  
45 variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 7. Preferred fragments of (b)  
comprise an epitope from SEQ ID NO: 7. Other preferred fragments lack one or more amino acids



(e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 7. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 7 is removed. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

**(5) GAS 040**

GAS 040 corresponds to M1 GenBank accession numbers GI:13621545 and GI:15674449, to M3 GenBank accession number GI: 21909733, to M18 GenBank accession number GI:19745402, and is also referred to as 'Spy0269' (M1), 'SpyM3\_0197' (M3), 'SpyM18\_0256' (M18) and 'prgA'. GAS 040 has also been identified as a putative surface exclusion protein. Amino acid and polynucleotide sequences of GAS 040 from an M1 strain are set forth below:

**SEQ ID NO: 9**

MDLEQTKPNQVKQKIALTSTIALLSASVGVSHQVKADDRASGETKASNTHDDSLPKPETIQEAKATIDAV  
BKTLSSQQAELTELALTKTTABINHLKEQQDNEQKALTSAQEIYTNLTASSEETLLAQGAHQRELT  
TETELHNAQADQHSKETALSEQKASI SAETTRAQDLVEQVKTSEQNIAKLNAMISNPDAITKAAQTANDN  
TKALSSELEKAKADLENQKAKVKKQLTEELAAQKAALAEKEABLSRLKSSAPSTQDSIVGNNTMKAPQGY  
PLBELKKLEASGYIGSASYNYYKEHADQIIAKASPGNQLNQYQDIPADRNRFVDPDNLTPEVQNBQAQF  
AAHMINSVRRQLGLPPVTVTAGSQEPARLLSTSYKKTGHNTGPSFVYGPVSGHYGVGPHDKTI IEDSA  
GASGLIRNDDNMYENIGAFNDVHTVNGIKRGIYDSIKYMLFTDHLHGNTYGHAINFLRVDKHNPNAPVYL  
GFSTSNVGSLEHFMFPESNIAHQRFNKTPIKAVGSTKDYAQRVGTVSDTIAAIKGVSSLENRLSAI  
HQBADIMAAQKVSQLOGKLASTLKQSDSLNLQVRQLNDTKGSLRTELLAAKAKQAQLEATRDQSLAKLA  
SLKAALHQTEALAEQAARVTALVAKKAHLOYLRDFKLNPNRLQVIRERIDNTKQDLAKTTSSLLNAQEA  
LAALQAKQSSLEATIATTEHQTLTKLANEKEYRHLDEDIATVPDLQVAPPLTGKPLSYSKIDTTPLV  
QEMVKETKQLEASARLAAENTSLVAEALVGQTSEMVASNAIVSKITSSITQPSKTSYSGSGSSTTSNLI  
SDVDESTQRALKAGVVMLAAVGLTGFRFRKESK

**SEQ ID NO: 10**

ATGGACTTAGAACAAACGAAGCCAAACCAAGTTAAGCAGAAAATTGCTTTAACCTCAACAATTGCTTTAT  
TGAGTGCCAGTGTAGGCGTATCTACCAAGTCAAAGCAGATGATAGAGCCTCAGGAGAAACGAAGGCGAG  
TAATACTCAGCAGATAGTTTACCAAAACCAGAAACAATTCAAGAGGCAAAGGCAACTATTGATGCAGTT  
GAAAAAATCTCAGTCAACAAAAGCAGAACTGACAGAGCTTGCTACCGCTCTGACAAAACTACTGCTG  
AAATCAACCACTTAAAGAGCAGCAAGATAATGAACAAAAGCTTTAACCTCTGCACAAGAAATTTACAC  
TAATACTCTTGCAAGTAGTGAGGAGACGCTATTAGCCCAAGGAGCCGAACATCAAAGAGAGTTAACAGCT  
ACTGAAACAGAGCTTCATAATGCTCAAGCAGATCAACATTCAAAGAGAGCTGCATTGTCAGAACAAAAG  
CTAGCATTTCAGCAGAACTACTCGAGCTCAAGATTTAGTGGAACAAGTCAAACGCTGTAACAAAATAT  
TGCTAAGCTCAATGCTATGATTAGCAATCCTGATGCTATCACTAAAGCAGCTCAAACGGCTAATGATAAT  
ACAAAAGCATTAAAGCTCAGAATTGGAGAAGGCTAAAGCTGACTTAGAAAATCAAAGCTAAAGTTAAAA  
AGCAATTGACTGAAGAGTTGGCAGCTCAGAAAGCTGCTCTAGCAGAAAAAGAGGCAGAACTTAGTCGTCT  
TAAATCCTCAGCTCCGTCTACTCAAGATAGCATTGTGGGTAATAATACCATGAAAGCACCAGCAAGGCTAT  
CCTCTTGAAGAACTTAAAAAATTAGAAGCTAGTGGTTATATTGGATCAGCTAGTTACAATAATTATTACA  
AAGAGCATGCAGATCAAATTATTGCCAAAGCTAGTCCAGGTAATCAATTAAATCAATACCAAGATATTCC  
AGCAGATCGTAATCGCTTTGTTGATCCCGATAATTTGACACCAGAAGTGCAAAATGAGCTAGCGCAGTTT  
GCAGCTCACATGATTAATAGTGTAAGAAGACAATTAGGTCTACCACCAGTTACTGTTACAGCAGGATCAC  
AAGAATTTGCAAGATTACTTAGTACCAGCTATAAGAAAACCTCATGGTAATACAAGACCATCATTTGTCTA  
CGGACAGCCAGGGGTATCAGGGCATTATGGTGTGGGCCTCATGATAAACTATTATTGAAGACTCTGCC  
GGAGCGTCAGGGCTCATTCGAAATGATGATAACATGTACGAGAATATCGGTGCTTTTAACGATGTGCATA  
CTGTGAATGGTATTAAACGTGGTATTTATGACAGTATCAAGTATATGCTCTTTACAGATCATTTACACGG  
AAATACATACGGCCATGCTATTAACTTTTACGTGTAGATAAACATAACCTAATGCGCCTGTTTACCTT  
GGATTTTCAACCAGCAATGTAGGATCTTTGAATGAACACTTTGTAATGTTTCCAGAGTCTAACATTGCTA  
ACCATCAACGCTTTAATAAGACCCCTATAAAAGCCGTTGGAAGTACAAAAGATTATGCCCAAAGAGTAGG  
CACTGTATCTGATACTATTGCAGCGATCAAAGGAAAGTAAGCTCATTAGAAAATCGTTTGTGCGCTATT  
CATCAAGAAGCTGATATTATGGCAGCCCAAGCTAAAGTAAGTCAACTTCAAGGTAAATTAGCAAGCACAC  
TTAAGCAGTCAGACAGCTTAAATCTCCAAGTGAGACAATTAAATGATACTAAAGGTTCTTTGAGAACAGA



ATTACTAGCAGCTAAAGCAAAACAAGCACAACCTCGAAGCTACTCGTGATCAATCATTAGCTAAGCTAGCA  
 TCGTTGAAAGCCGCACTGCACCAGACAGAAGCCTTAGCAGAGCAAGCCGAGCCAGAGTGACAGCACTGG  
 TGGCTAAAAAAGCTCATTGCAATATCTAAGGGACTTTAAATTGAATCCTAACCGCCTTCAAGTGATACG  
 TGAGCGCATTGATAATACTAAGCAAGATTTGGCTAAAACTACCTCATCTTTGTTAAATGCACAAGAAGCT  
 5 TTAGCAGCCTTACAAGCTAAACAAAGCAGTCTAGAAGCTACTATTGCTACCACAGAACACCAGTTGACTT  
 TGCTTAAACCTTAGCTAACGAAAAGGAATATCGCCACTTAGACGAAGATATAGCTACTGTGCCTGATTT  
 GCAAGTAGCTCCACCTCTTACGGGCGTAAACCGCTATCATATAGTAAGATAGATACTACTCCGCTTGTT  
 CAAGAAATGGTTAAAGAAACGAAACAACCTATTAGAAGCTTCAGCAAGATTAGCTGCTGAAAATACAAGTC  
 TTGTAGCAGAAGCGCTTGTGGCCAAACCTCTGAAATGGTAGCAAGTAATGCCATTGTGTCTAAAATCAC  
 10 ATCTTCGATTACTCAGCCCTCATCTAAGACATCTTATGGCTCAGGATCTTCTACAACGAGCAATCTCATT  
 TCTGATGTTGATGAAAGTACTCAAAGAGCTCTTAAAGCAGGAGTCGTCATGTTGGCAGCTGTCGGCCTCA  
CAGGATTTAGGTTCCGTAAGGAATCTAAGTGA

Preferred GAS 040 proteins for use with the invention comprise an amino acid sequence: (a) having  
 15 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 9; and/or (b) which is a fragment of at least  $n$   
 consecutive amino acids of SEQ ID NO: 9, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 or more). These GAS 040 proteins include variants  
 (e.g. allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 9. Preferred  
 20 fragments of (b) comprise an epitope from SEQ ID NO: 9. Other preferred fragments lack one or  
 more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one  
 or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ  
 ID NO: 9. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of  
 SEQ ID NO: 9 is removed. As another example, in one embodiment, the underlined amino acid  
 25 sequence at the C-terminus of SEQ ID NO: 9 is removed. Other fragments omit one or more domains  
 of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane  
 domain, or of an extracellular domain).

#### (6) GAS 389

GAS 389 corresponds to M1 GenBank accession numbers GI:13622996 and GI:15675772, to M3  
 30 GenBank accession number GI: 21911237, to M18 GenBank accession number GI: 19746884, and is  
 also referred to as 'Spy1981' (M1), 'SpyM3\_1701' (M3), 'SpyM18\_2045' (M18) and 'relA'. GAS  
 389 has also been identified as a (p)ppGpp synthetase. Amino acid and polynucleotide sequences of  
 GAS 389 from an M1 strain are set forth below:

#### SEQ ID NO: 11

35 MRNEMAKIMNVTGEEVIALAATYMTKADVAFVAKALAYATAAHFYQVRKSGEPYIVHPPIQVAGILADLHL  
 DAVTVACGFLHDVVEDTDITLDEIRADFGHDARDIVDGVTKLGEVEYKSHEEQLAENHRKMLMAMSKDIR  
 VILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGISRIKWELEDLAFRYLNETEFYKISHM  
 MKEKRREREALVEAIVSKVKTYTTQOGLFGDVYGRPKHIYSIYRKMDDKKRFDQIFDLIAIRCVMETQS  
 DVYAMVGYIHELWRPMPGRPKDYIAAPKANGYQSIHTTVYGPKGPIEQIRTKDMHQAIEYGVAAHWAYK  
 40 KGVRGKVNQAEQAVGMNWIKEVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKESGPIDF  
 AYAIHTQIGEKATGAKVNGRMVPLTAKLKTGDVVEIITNANSFGPSRDWVKLVKTNKARNKIRQFFKNQD  
 KELSVNKGRLDLVSFYQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFGDISPISVFNKLTEKER  
 REEBRAKAKAEAEELVKGGEVKHENKDVVKVRSNGVI IQGASGLLMRIAKCCNPVPGDPIDGYITKGRG  
 45 IAIHRSDCHNIKSQDGYQERLIEVEWDLNSSKDYQAEIDYGLNRSGLLNDVLQILSNSTKSISTVNAQ  
 PTKDMKFANIHVSFGIPNLTHLTTVVEKIKAVPDVYSVKRTNG

#### SEQ ID NO: 12

ATGAGGAACGAAATGGCAAAAATAATGAACGTAACAGGAGAAGAAGTCATTGCCTTAGCGGCCACCTATA



TGACCAAGGCTGATGTGGCTTTTGTGGCAAAGGCTTTAGCATATGCAACAGCGGCCCATTTCTACCAAGT  
 GAGAAAGTCAGGCGAACCCCTATATCGTCCATCCGATTCAGGTGGCGGGGATTCTGGCTGATTGTCATCTG  
 GATGCTGTGACAGTTGCTTGTGGCTTTTACATGATGTCTAGAGAATACGGATATTACCTTAGATGAGA  
 TCGAAGCAGACTTTGGCCATGATGCTCGTGATATCGTTGATGGTGTACCAAGTTAGGTGAAGTTGAGTA  
 5 CAAATCTCATGAGGAGCAACTCGCCGAAAACCATCGCAAAATGCTGATGGCTATGTCCAAAGATATTGCG  
 GTGATTTTGGTGAAATTTGGCTGACCGCCTGCATAATATGCGCACCCCTCAAACATTTGCGCAAGGACAAAC  
 AAGAGCGCATTTTCGCGCGAAACCATGGAAATCTATGCCCCCTTGGCGCATCGTTTGGGGATTAGTCGCAT  
 CAAATGGGAACTAGAAGATTGGCTTTTCGTTACCTCAATGAAACCGAATTTTACAAAATTTCCCATATG  
 10 ATGAAAGAAAAACGTCGCGAGCGTGAAAGCTTTGGTAGAGGCTATTGTGAGTAAGGTCAAAACCTATACGA  
 CACAACAAGGGTTGTTTGGAGATGTGTATGGCCGACCAAAACACATTTATTGATTTATCGGAAAATGCG  
 GGACAAAAGAAACGATTTCGATCAGATTTTGTATCTGATTGCCATTCTGTTGTGTCATGGAAACGCAAAGC  
 GATGTCTATGCTATGGTTGGCTATATTGATGAGCTTTGGCGTCCATGCCAGGCCGCTTCAAGGATTATA  
 TTGCAGCTCCTAAAGCTAATGGCTACCAGTCTATTCATACCACCGTGTATGGGCCAAAAGGACCTATTGA  
 GATTCAAATCAGAACTAAGGACATGCATCAAGTGGCTGAGTACGGGGTTGCTGCTCACTGGGCTTATAAA  
 15 AAAGGCGTGCGTGGTAAGGTCAATCAAGCTGAGCAAGCCGTTGGCATGAACTGGATCAAAGAGCTGGTAG  
 AATTGCAAGATGCCTCAAATGGCGATGCAGTGGACTTTGTGGATTTCGGTCAAAGAAGACATTTTTTCTGA  
 ACGGATTTATGTCTTTACACCGACAGGGGCGGTTTCAGGAGTTACCAAAGAATCAGGTCTTATTGATTTT  
 GCTTATGCGATCCATACGCAAAATCGGTGAAAAGCAACAGGTGCCAAAGTCAATGGACGTATGGTTCCTC  
 TCACTGCCAAGTTAAAAACAGGAGATGTGGTTGAAATCATCACCAATGCCAATTCCTTTGGCCCTAGTCG  
 20 AGACTGGGTAAACTGGTCAAACCAATAAGGCTCGCAACAAAATTCGTCAGTTCTTTAAAAATCAAGAC  
 AAGGAATTGTCAGTGAATAAAGGCCGTGATTTGTGGTGTCTTATTTTCAAGAGCAGGGCTACGTTGCCA  
 ATAAATACCTTGACAAAAACGCATTGAAGCCATCCTTCCAAAAGTCAGTGTGAAGAGCGAAGAATCACT  
 CTATGCAGCCGTTGGGTTTGGTGACATTAGTCCTATCAGTGTCTTTAACAAGTTAACCGAAAAAGAGCGC  
 CGTGAAGAAGAAAGGGCCAAGGCTAAAGCAGAAGCTGAAAGATTGGTTAAGGGCGGTGAGGTCAAACACG  
 25 AAAACAAAGATGTGCTCAAGGTTCCGAGTGAAGATGGAGTCATTATCCAAGGAGCATCAGGCCTCTTGAT  
 GCGGATTGCCAAGTGTTGTAATCCTGTACCTGGTGATCCTATTGACGGCTACATTACCAAAGGGCGTGGC  
 ATTGCGATTACAGATCGGACTGTACATAACATTAAGAGTCAAGATGGCTACCAAGAACGCTTGATTGAGG  
 TCGAGTGGGATTTGGACAATTCGAGTAAAGATTATCAGGCTGAAATTGATATCTATGGGCTCAATCGTAG  
 TGGTCTGCTTAATGATGTGCTCCAAATTTTATCAAACCAAGAGCATATCGACAGTCAATGCTCAG  
 30 CCGACCAAGGACATGAAGTTTGCTAATATTACGTTGAGCTTTGGCATTCCAAATCTGACGCATCTGACCA  
 CTGTTGTGCAAAAAATCAAGGCAGTTCCAGATGTTTATAGCGTGAAGCGGACCAATGGCTAA

Preferred GAS 389 proteins for use with the invention comprise an amino acid sequence: (a) having  
 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
 35 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 11; and/or (b) which is a fragment of at least  $n$   
 consecutive amino acids of SEQ ID NO: 11, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 or more). These GAS 389 proteins include variants  
 (e.g. allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 11. Preferred  
 fragments of (b) comprise an epitope from SEQ ID NO: 11. Other preferred fragments lack one or  
 40 more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one  
 or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ  
 ID NO: 11. Other fragments omit one or more domains of the protein (e.g. omission of a signal  
 peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

#### (7) GAS 504

45 GAS 504 corresponds to M1 GenBank accession numbers GI:13622806 and GI:15675600, to M3  
 GenBank accession number GI: 21911061, to M18 GenBank accession number GI: 19746708, and is  
 also referred to as 'Spy1751' (M1), 'SpyM3\_1525', 'SpyM18\_1823' (M18) and 'fabK'. GAS 504  
 has also been identified as a putative trans-2-enoyl-ACP reductase II. Amino acid and polynucleotide  
 sequences of GAS 504 of an M1 strain are set forth below:

50 SEQ ID NO: 13



MKTRITELLNIDYPIFOGGMAWVADGDLGAVSNAGGLQIIGGGNAPKEVVKANIDRVKAITDRPFGVNI  
 MLLSPFADDIVDLVIEGVKVVTGAGNPGKYMERLHQAGIIVVPVPSVALAKRMEKLGVDVIAEGME  
 AGGHIGKLTMSLVRQVVEAVSIPVIAAGGIADGHGAAAFMLGABAVQIGTRFVAKESNAHQNFKDKI  
 LAAKDIDTVISAQVVGHVPVRSIKNKLTSAYAKAEKAPFLIGQKTATDIBENGAGSLRHAVIEGDVVNGSVM  
 5 AGQIAGLVRKEESCETILKDIYYGAARVIOEAKRWQSVSIBK

**SEQ ID NO: 14**

ATGAAAACACGTATTACAGAATTACTTAATATTGATTACCCCATTTTCAAGGAGGAATGGCTTGGGTTG  
 CTGATGGTGATTTAGCAGGTGCAGTTTCTAATGCTGGTGGTTAGGCATTATAGGTGGTGGCAATGCTCC  
 10 CAAAGAAGTCGTTAAAGCTAATATTGATCGTGTCAAAGCTATTACTGATAGACCTTTTGGGGTTAATATC  
 ATGCTTTTATCTCCTTTTGTGATGATATCGTTGATCTGGTCATTGAAGAAGGTGTTAAAGTAGTAACAA  
 CAGGCGCAGGAAATCCAGGAAAGTATATGGAAAGACTGCACCAGGCGGTATAATCGTTGTTCTGTGTG  
 CCCAAGCGTTGCGCTAGCCAAACGTATGGAAAGCTTGGGGTAGATGCTGTTATTGCTGAGGGTATGGAA  
 GCTGGAGGACATATTGGCAAGTTAACGACTATGTCTTTAGTAAGACAAGTTGTTGAAGCGGTTTCGATTC  
 15 CTGTCATTGCGGCAGGTGGTATAGCTGATGGTCATGGTGCAGCAGCAGCATTATGTTAGGAGCAGAGGC  
 TGTTCAAATTGGAACCTCGCTTGTGTTGTTGCTAAAGAATCCAATGCTCACCAAATTTTAAAGATAAAATC  
 TTAGCAGCAAAAGATATTGATACGGTGATTTCTGCGCAGGTTGTGGGCCACCTGTCCGTTCTATTAAAA  
 ATAAATTGACCTCAGCTTACGCTAAAGCAGAAAAGCATTTTTAATTGGTCAAAAACAGCTACTGATAT  
 TGAAGAAATGGGAGCAGGATCGCTTCGACACGCTGTTATTGAAGGCGATGTAGTCAATGGATCTGTTATG  
 20 GCTGGCCAAATTGCAGGGCTTGTGAGAAAAGAAGAAAGCTGTGAAACGATTTTAAAGATATTATTATG  
 GTGCAGCTCGTGTTATTCAAAATGAAGCTAAGCGCTGGCAATCTGTTTCAATAGAAAAGTAG

Preferred GAS 504 proteins for use with the invention comprise an amino acid sequence: (a) having  
 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
 25 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 13; and/or (b) which is a fragment of at least  $n$   
 consecutive amino acids of SEQ ID NO: 13, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
 30, 35, 40, 50, 60, 70, 80, 90, 100, 150 or more). These GAS 504 proteins include variants (e.g. allelic  
 variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 13. Preferred fragments of (b)  
 comprise an epitope from SEQ ID NO: 13. Other preferred fragments lack one or more amino acids  
 30 (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino  
 acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 13.  
 Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a  
 cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

**(8) GAS 509**

35 GAS 509 corresponds to M1 GenBank accession numbers GI:13622692 and GI:15675496, to M3  
 GenBank accession number GI: 21910899, to M18 GenBank accession number GI: 19746544, and is  
 also referred to as 'Spy1618' (M1), 'SpyM3\_1363' (M3), 'SpyM18\_1627' (M18) and 'cysM'. GAS  
 509 has also been identified as a putative O-acetylserine lyase. Amino acid and polynucleotide  
 sequences of GAS 509 of an M1 strain are set forth below:

**SEQ ID NO: 15**

MTKIYKTITELVGQTPIIKLNRLIPNEAADVYVKLEAFNPGSSVKDRIALSMIEAAEAEGLISPGDVIIIE  
 PTSGNTGIGLAWVGAAGYRVIIIVMPETMSLERRQIIQAYGAELVLTPGAEGMKGAIAKAETLAIELGAW  
 MPMQFNNPANPSIHEKTTAQEILEAFKEISLDAFVSGVGTGGTSLSGVSHVLKKANPETVIYAVEABESAV  
 LSGQEPGPHKIQGISAGFIPNTLDTKAYDQIIRVKS KDALETARLTGAKEGFLVGISSGAALYAAIEVAK  
 45 QLGKGKHVLTILPDNGERYLSTELYDVPVIKTK

**SEQ ID NO: 16**

ATGACTAAAATTTACAAAACCTATAACAGAATTAGTAGGTCAAACACCTATTATCAAACCTTAACCGTTTAA  
 TTCAAACGAAGCTGCTGACGTTTATGTAAAATTAGAAGCTTTTAAACCCAGGATCTTCTGTAAAGATCG  
 50 TATTGCTTTATCGATGATTGAAGCTGCTGAAGCTGAAGGTCTGATAAGTCCTGGTGACGTTATTATCGAA



CCAACAAGTGGAATACAGGTATTGGTCTTGCATGGGTAGGTGCTGCTAAAGGGTATCGAGTCATTATTG  
 TTATGCCCGAAACTATGAGCTTGGAAGACGGCAATCATTGAGGCTTATGGTGCAGAGCTTGTCTTAAC  
 ACCTGGAGCAGAAGGTATGAAAGGGGCTATTGCAAAAGCTGAAACTTTAGCAATAGAAGTAGGTGCTTGG  
 ATGCCTATGCAATTTAATAACCCTGCCAATCCAAGCATCCATGAAAAACAACAGCTCAAGAAATTTTGG  
 5 AAGCTTTTAAGGAGATTTCTTTAGATGCATTTCGTATCTGGTGTGGTACTGGAGGAACACTTTCTGGTGT  
 TTCACATGTCTTGAAAAAGCTAACCCGTGAACTGTTATCTATGCTGTTGAAGCTGAAGAATCTGCTGTC  
 TTATCTGGTCAAGAGCCTGGACCACATAAAATTCAAGGTATATCAGCTGGATTATCCCAAACACGTTAG  
 ATACCAAAGCCTATGACCAAATTATCCGTGTTAAATCGAAAGATGCTTTAGAACTGCTCGACTAACAGG  
 AGCTAAGGAAGGCTTCTCTGGTTGGGATTTCTTCTGGAGCTGCTCTTTACGCCGCTATTGAAGTCGCTAAA  
 10 CAGTTAGGAAAAGGCAAACATGTGTTAACTATTTTACCAGATAATGGCGAACGCTATTTATCGACTGAAC  
TCTATGATGTACCAGTAATTAAGACGAAATAA

Preferred GAS 509 proteins for use with the invention comprise an amino acid sequence: (a) having  
 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
 15 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 15; and/or (b) which is a fragment of at least  $n$   
 consecutive amino acids of SEQ ID NO: 15, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
 30, 35, 40, 50, 60, 70, 80, 90, 100, or more). These GAS 509 proteins include variants (e.g. allelic  
 variants, homologs, orthologs, paralog, mutants, etc.) of SEQ ID NO: 15. Preferred fragments of (b)  
 comprise an epitope from SEQ ID NO: 15. Other preferred fragments lack one or more amino acids  
 20 (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino  
 acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 15. For  
 example, in one embodiment, the underlined amino acid sequence at the C-terminus of SEQ ID NO:  
 15 is removed. Other fragments omit one or more domains of the protein (e.g. omission of a signal  
 peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

## 25 (9) GAS 366

GAS 366 corresponds to M1 GenBank accession numbers GI:13622612, GI:15675424 and  
 GI:30315979, to M3 GenBank accession number GI: 21910712, to M18 GenBank accession number  
 GI: 19746474, and is also referred to as 'Spy1525' (M1), 'SpyM3\_1176' (M3), 'SpyM18\_1542'  
 (M18) and 'murD'. GAS 366 has also been identified as a UDP-N-acetylemuramoylalanine-D-  
 30 glutamate ligase or a D-glutamic acid adding enzyme. Amino acid and polynucleotide sequences of  
 GAS 366 of an M1 strain are set forth below:

### SEQ ID NO: 17

MKVISNFQNKKILILGLAKSGEAAKLLTKLGALVTVNDSPFDQNPAAQALLEEGIKVICGSHPVVELLD  
 ENFEYMKVKNPGIPYDNPMVKRALAKEIPILTEVELAYFVSEAPIIGITGSNGKTTTTMTIADVLNAGGQS  
 35 ALLSGNIGYPASKVVQKAIAGDTLVMELSSFQLVGVNAFRPHIAVITNLMPTHLDYHGSFEDYVAAKWMI  
 QAQMTESDYLIILNANQEISATLAKTTKATVIPFSTQKVVDGAYLKDGILYFKEQAI AATDLGVPGSHNI  
 ENALATI AVAKLSGIADDIAQCLSHFGGVKHLRQVQIKDITFYNDKSTNIIATQKALSGFDNSRLI  
 LIAGGLDRGNEFDDLVPDLLGLKQMI ILGESABRMKRAANKAEVSYLEARNVAEATELAFKLAQTGDTIL  
 LSPANASWDMYPNFEVRGDEFLATFDCLRGDA

### 40 SEQ ID NO: 18

ATGAAAGTGATAAGTAATTTTCAAAACAAAAAATATTAATATTGGGGTTAGCCAAATCGGGCGAAGCAG  
 CAGCAAATTTATTGACCAAACCTTGGTGCTTTAGTGACTGTTAATGATAGTAAACCATTTGACCAAATCC  
 AGCGGCACAAGCCTTGTTGGAAGAGGGGATTAAGGTCATTTGTGGTAGCCACCCAGTAGAATTATTAGAT  
 GAGAACTTTGAGTACATGGTTAAAAACCTGGGATTCCTTATGATAATCCTATGGTTAAACGCGCCCTTG  
 45 CAAAGGAAATTTCCCATCTTGACTGAAGTAGAATTGGCTTATTTTCGTATCTGAAGCGCCTATTATCGGGAT  
 TACAGGATCAAACGGGAAGACAACCACAACGACAATGATTGCCGATGTTTTGAATGCTGGCGGGCAATCT  
 GCACTCTTATCTGGAAACATTGGTTATCCTGCTTCAAAAGTTGTTCAAAAAGCAATTGCTGGTGATACTT  
 TGGTGATGGAATTGTCCTCTTTTCAATTAGTGGGAGTGAATGCTTTTCGCCCTCATATTGCTGTCATCAC



TAATTTAATGCCGACTCACCTGGACTATCATGGCAGTTTTGAGGATTATGTTGCTGCTAAATGGATGATT  
CAAGCTCAGATGACAGAATCAGACTACCTTATTTTAAATGCTAATCAAGAGATTCAGCAACTCTAGCTA  
AGACCACCAAAGCAACAGTGATTCTTTTCAACTCAAAAAGTGGTTGATGGAGCTTATCTGAAGGATGG  
AATACTCTATTTTAAAGAACAGGCGATTATAGCTGCAACTGACTTAGGTGTCCAGGTAGCCACAACATT  
5 GAAAATGCCCTAGCAACTATTGCAGTTGCCAAGTTATCTGGTATTGCTGATGATATTATTGCCCAGTGCC  
TTTCACATTTTGGAGGCGTTAAACATCGTTTGCAACGGGTTGGTCAAATCAAAGATATTACCTTCTACAA  
TGACAGTAAGTCAACCAATATTTTAGCCACTCAAAAAGCTTTATCAGGTTTTGATAACAGTCGCTTGATT  
TTGATTGCTGGCGGTCTAGATCGTGGCAATGAATTTGACGATTTGGTGCCAGACCTTTTAGGACTTAAGC  
AGATGATTATTTTGGGAGAATCCGCAGAGCGTATGAAGCGAGCTGCTAACAAAGCAGAGGTCTCTTATCT  
10 TGAAGCTAGAAATGTGGCAGAAGCAACAGAGCTTGCTTTTAAAGCTGGCCCAAACAGGCGATACTATCTTG  
CTTAGCCCAGCCAATGCTAGCTGGGATATGTATCCTAATTTTGAGGTTCTGTGGGATGAATTTTGGCAA  
CCTTTGATTGTTTAAAGAGGAGATGCCTAA

Preferred GAS 366 proteins for use with the invention comprise an amino acid sequence: (a) having  
15 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 17; and/or (b) which is a fragment of at least  $n$   
consecutive amino acids of SEQ ID NO: 17, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
30, 35, 40, 50, 60, 70, 80, 90, 100, 150 or more). These GAS 366 proteins include variants (e.g. allelic  
variants, homologs, orthologs, paralog, mutants, etc.) of SEQ ID NO: 17. Preferred fragments of (b)  
20 comprise an epitope from SEQ ID NO: 17. Other preferred fragments lack one or more amino acids  
(e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino  
acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 17. For  
example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO:  
17 is removed. Other fragments omit one or more domains of the protein (e.g. omission of a signal  
25 peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

#### (10) GAS 159

GAS 159 corresponds to M1 GenBank accession numbers GI:13622244 and GI:15675088, to M3  
GenBank accession number GI: 21910303, to M18 GenBank accession number GI: 19746056, and is  
also referred to as 'Spy1105' (M1), 'SpyM3\_0767' (M3), 'SpyM18\_1067' (M18) and 'potD'. GAS  
30 159 has also been identified as a putative spermidine/putrescine ABC transporter (a periplasmic  
transport protein). Amino acid and polynucleotide sequences of GAS 159 of an M1 strain are set  
forth below:

#### SEQ ID NO: 19

MRKLYSFLAGVLGVIVILTSLSFILQKKSGSGSQSDKLVIYNWGDYIDPALLKKFTKETGIEVQYETFDS  
35 NEAMYTKIKQGGTTYDIAVPSDYTIDKMIKENLLNKLDSKLVGMDNIGKEFLGKSFPQNDYSLPYFWG  
TVGIVYNDQLVDKAPMHWEDLWRPEYKNSIMLIDGAREMLGVGLTTFGYSVNSKNLEQLQAAERKLOQLT  
PNVKAI VADEMKGYMIQGDAAGITFSGEASEMLDSNEHLHYIVPSEGSNLWFDNLVLPKTMKHEKEAYA  
FLNFINRPENAAQNAAYIGYATPNKKAKALLPDEIKNDPAFYPTDDI I KKLEVYDNLGSRWLGIYNDLYL  
40 QFKMYRK

#### SEQ ID NO: 20

ATGCGTAAACTTTATTCTTTCTAGCAGGAGTTTTGGGTGTTATTGTTATTTTAAACAAGTCTTTCTTTCA  
TCTTGCGAGAAAAATCGGGTTCTGGTAGTCAATCGGATAAATTAGTTATTTATAACTGGGGAGATTACAT  
TGATCCAGCTTTGCTCAAAAATTACCAAAGAAACGGGCATTGAAGTGCAGTATGAAACTTTGATTCC  
45 AATGAAGCCATGTACACTAAAATCAAGCAGGCGGAACCACTTACGACATTGCTGTTCTTAGTGATTACA  
CCATTGATAAAATGATCAAAGAAAACCTACTCAATAAGCTTGATAAGTCAAATTAGTTGGCATGGATAA  
TATCGGGAAGAATTTTATAGGGAAAAGCTTTGACCCACAAAACGACTATTCTTTGCCTTATTTCTGGGGA  
ACCGTTGGGATTGTTTATAATGATCAATTAGTTGATAAGGCGCTATGCACTGGGAAGATCTGTGGCGTC  
CAGAATATAAAAATAGTATTATGCTGATTGATGGAGCGCGTGAAATGCTAGGGGTTGGTTTAAACAATTT



TGGTTATAGTGTGAATTCTAAAAATCTAGAGCAGTTGCAGGCAGCCGAGAGAAAACCTGCAGCAGTTGACG  
CCGAATGTTAAAGCCATTGTAGCAGATGAGATGAAAGGCTACATGATTCAAGGTGACGCTGCTATTGGAA  
TTACCTTTTCTGGTGAAGCCAGTGAGATGTTAGATAGTAACGAACACCTTCACTACATCGTGCCTTCAGA  
AGGGTCTAACCTTTGGTTTGATAATTTGGTACTACCAAAAACCATGAAACACGAAAAAGAAGCTTATGCT  
5 TTTTGAACCTTATCAATCGTCCTGAAAATGCTGCGCAAAATGCTGCATATATTGGTTATGCGACACCAA  
ATAAAAAAGCCAAGGCCTTACTTCCAGATGAGATAAAAAATGATCCTGCTTTTATCCAACAGATGACAT  
TATCAAAAAATTGGAAGTTTATGACAATTTAGGGTCAAGATGGTTGGGGATTATAATGATTATACCTC  
CAATTTAAATGTATCGCAAATAA

- 10 Preferred GAS 159 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 19; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 19, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150 or more). These GAS 159 proteins include variants (e.g. allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 19. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 19. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 19. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 19 is removed. In another example, the underlined amino acid sequence at the C-terminus of SEQ ID NO: 19 is removed. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

#### (11) GAS 217

- GAS 217 corresponds to M1 GenBank accession numbers GI:13622089 and GI:15674945, to M3 GenBank accession number GI: 21910174, to M18 GenBank accession number GI: 19745987, and is also referred to as 'Spy0925' (M1), 'SpyM3\_0638' (M3), and 'SpyM18\_0982' (M18). GAS 217 has also been identified as a putative oxidoreductase. Amino acid and polynucleotide sequences of GAS 217 of an M1 strain are set forth below:

#### SEQ ID NO: 21

- 30 MAQRIVITGASGGLAQAIKQLPKEDSLILLGRNKRLEHCYQHIDNKECLELDITNPVAIEKMQVAQIY  
QRYGRIDVLINNAGYGAFKGFEEFSAQBIADMFQVNTLASIHFAFLIGQKMAEQGQHLINIVSMAGLIA  
SAKSSIYSATKFALIGFSNALRLELADKGVYVTTVNP GPIATKFPDQADPSGHYLESVGKFTLQPNQVAK  
RLVSIIGKNKRELNL PFS LAVTHQFYTLFPKLS DY LARKVFNKY

#### SEQ ID NO: 22

- ATGGCACAAAGAATCATTGTTATCACGGGAGCTTCTGGAGGACTGGCTCAGGCAATTGTTAAGCAGTTAC  
CCAAGGAAGACAGCTTGATTTTATTAGGACGTAACAAAGAACGCCTAGAACACTGTTATCAGCATATTGA  
CAACAAAGAATGCCTCGAGTTGGATATTACCAATCCAGTAGCCATTGAGAAAATGGTCGCCCAGATTAC  
40 CAGCGCTATGGCCGTATTGATGTCTTGATTAATAATGCTGGCTACGGAGCTTCAAAGGCTTTGAAGAGT  
TTTCTGCCCAAGAAATAGCTGATATGTTTCAGGTTAACACCCTAGCGAGCATTCACTTTGCTTGCTTGAT  
TGGTCAGAAAATGGCAGAGCAGGGGCAAGGTCACCTTATTAATATTGTGTCCATGGCAGGCTTGATTGCG  
TCAGCCAAATCGAGCATTTATTCAGCCACCAAGTTTGCCCTTATCGGATTTTCCAATGCCCTTCGCTTAG  
AATTAGCGGATAAAGGGGTTTACGTGACCACCGTGAATCCAGGTCCCATTGCCACCAAGTTTTTTGACCA  
45 AGCTGACCCGTCTGGACATTATTGGAAGCGTTGGTAAATTTACTCTCCAACCAATCAAGTGGCTAAG  
CGTTTGGTTTCTATTATCGGGAAAAATAACGAGAATTGAATTTGCCCTTTAGTTTAGCGGTGACCCATC  
AATTTTACACCCTTTTCCCTAAATTATCTGATTATCTTGCAAGAAAGGTATTTAATTATAAATGA



Preferred GAS 217 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 21; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 21, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, or more). These GAS 217 proteins include variants (e.g. allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 21. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 21. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 21. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

### (12) GAS 309

GAS 309 corresponds to M1 GenBank accession numbers GI:13621426 and GI:15674341, to M3 GenBank accession number GI: 21909633, to M18 GenBank accession number GI: 19745363, and is also referred to as 'Spy0124' (M1), 'SpyM3\_0097' (M3), 'SpyM18\_0205' (M18), 'nra' and 'rofA'. GAS 309 has also been identified as a regulatory protein and a negative transcriptional regulator. Amino acid and polynucleotide sequences of GAS 309 of an M1 strain are set forth below:

#### SEQ ID NO: 23

MIEKYLESSIESKCQLIVLFFKTSYLPITEVAEKTGLTFLQLNHYCEELNAFFPGSLSMITQKRMISCQF  
THPFKETYLYQLYASSNVLQLLAFLIKNGSHSRPLTDFARSHFLSNSSAYRMREALIPLLRNFELKLSKN  
KIVGBEYRIRYLIALLYSKPGIKVYDLTQDQNTIHSFLSHSSTHLKTSPLWSESFSFYDILLALSWKRH  
QFSVTIPQTRIFQQLKKLFVYDSLKKSSHDIIETYCQLNFSAGDLDYLYLIYITANNSFASLQWTPEHIR  
QYCQLFBENDTFRLLLNPIITLLPNLKEQKASLVKALMFPSKSPFLNLQHFIPETNLFVSPYYKGNQKLY  
TSLKLIIVEEWMKLPGRDLNKHKHFHLFCHYVEQSLRNIQPPLVVVFVASFNAHLLTDSFPRYFSDKS  
IDFHSYLLQDNVYQIPDLKPDLVITHSQLIPFVHHELTGKIABISFDESILSIQELMYQVKEEFQA  
DLTKQLT

#### SEQ ID NO: 24

TTGATAGAAAAATACTTGAATCATCAATCGAATCAAAATGTCAGTTAATTGTCTTGTTTTTTAAGACAT  
CTTATTGCAATAACTGAGGTAGCAGAAAAAAGTGGCTTAACCTTTTTACAATAAACCATTATTGTGA  
GGAAGTGAATGCCCTTTTCCCTGGTAGTCTGTCTATGACCATCCAAAAAGGATGATATCTTGCCAATTT  
ACACATCCTTTTAAAGAACTTATCTTTACCAACTCTATGCATCATCTAATGTCTTACAATTACTAGCCT  
TTTTAATAAAAAATGGTTCCCACTCTCGTCCCCTTACGGATTTTGCAAGAAGTCATTTTTATCAAATC  
CTCAGCTTATCGGATGCGCGAAGCATTGATTCTTTATTAAGAACTTTGAATTAAACTCTCTAAGAAC  
AAGATTGTCCGTGAGGAATATCGCATCCGTACCTCATCGCTCTGTCTATATAGTAAGTTTGGCATTAAAG  
TTTATGACTTGACGCAGCAAGACAAAAACACTATTCATAGCTTTTATCCCATAGTTCACCCACCTTAA  
AACCTCTCCTTGTTATCGGAATCGTTTTCTTTCTATGACATTTTATTAGCTTTATCGTGGAAGCGGCAT  
CAATTTTCGGTAATATTCCCCAAACCAGAAATTTTCAACAATTAAAAAACTTTTGTCTACGATTCTT  
TGAAAAAAGTAGCCATGATATTATCGAACTTACTGCCAACTAACTTTTACAGCAGGAGATTGGACTA  
CCTCTATTTAATTTATATCACCGCTAATAATCTTTTGCGAGCTTACAATGGACACCTGAGCATATCAGA  
CAATATTGTCAACTTTTTGAAGAAATGATACTTTTCGCCTGCTTTTAAATCCTATCATCACTCTTTTAC  
CTAACCTAAAGAGCAAAAGGCTAGTTTAGTAAAGCTCTTATGTTTTTTTCAAAATCATTCTTGTTTAA  
TCTGCAACATTTTATTCTGAGACCAACTTATTCGTTTCTCCGTACTATAAAGGAAACCAAAACTCTAT  
ACGTCCTTAAAGTTAATTGTCGAAGAGTGGATGGCCAACTTCTGGTAAGCGTGACTTGAACCATAAGC  
ATTTTCATCTTTTTTGCCACTATGTCGAGCAAAGTCTAAGAAATATCCAACCTCCTTTAGTTGTTGTTTT  
CGTAGCCAGTAATTTTATCAATGCTCATCTCCTAACGGATTCTTTTCCAAGGTATTTCTCGGATAAAAGC  
ATTGATTTTCATTCCTATTATCTATTGCAAGATAATGTTTATCAAATTCCTGATTTAAAGCCAGATTGG  
TCATCACTCACAGTCAACTGATTCTTTTGTTCACCATGAACCTACAAAAGGAATTGCTGTTGCTGAAAT  
ATCTTTTGATGAATCGATTCTGTCTATCCAAGAATTGATGTATCAAGTTAAAGAGGAAAAATTCCAAGCT  
GATTTAACCAAGCAATTAACATAA



Preferred GAS 309 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 23; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 23, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, or more). These GAS 309 proteins include variants (e.g. allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 23. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 23. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 23. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

**(13) GAS 372**

GAS 372 corresponds to M1 GenBank accession numbers GI:13622698 and GI:15675501, to M3 GenBank accession number GI: 21910905, to M18 GenBank accession number GI: 19746500 and is also referred to as 'Spy1625' (M1), 'SpyM3\_1369' (M3), and 'SpyM18\_1634' (M18). GAS 372 has also been identified as a putative protein kinase or a putative eukaryotic-type serine/threonine kinase. Amino acid and polynucleotide sequences of GAS 372 of an M1 strain are set forth below:

**SEQ ID NO: 25**

MIQIGKLFAGRYRILKSI GRGGMADV LANDLI LDNEDVAIKVLRTNYQTDQVAVARFQREBARAMAELNH  
PNIVAIRDIGBEDGQQFLVMEYVDGADLKRYIQNHAPLSNNEVVRIMEEVL SAMTLAHQKGI VHRDLKPO  
NILLTKEGVVKTDFGIAVAF AETS LTQTNSMLG SVHYLSPEQARGSKATI QSDIYAMGIMLFEMLTGHI  
PYDGDSAVTIALQHFQKPLPSII EENHNVPQALENVVIRATAKKLS DRYGSTFEMSRDLMTALSYNRSRE  
RKIIFENVESTKPLPKVASGPTASVKLS PPTPTVL TQESRLDQTNQTDALQPPTKKKKSGRFLGTLFKIL  
PSFFIVGVALFTYLILTKPTS VKVPNVAGTSLKVAKQELYDVGLKVGKIRQIESDTVAEGNVVRTDPKAG  
TAKRQGSSITLYVSIGNKGFDMENYKGLDYQBAMNSLIETYGVPSKIKIERIVTNEY PENTVISQSPSA  
GDKFNPNGKSKITLSVAVSDTITMPMVTBYSYADAVNTLTALGIDASRIKAYVPSSSSATGFVPIHSPSS  
KAIVSGQSPYYGTSLSLSDKGEISLYLYPEETHSSSSSSSSSTSSSNSSSINDSTAPGSNTLSPSETTSQ  
TP

**SEQ ID NO: 26**

ATGATTCAGATTGGCAAATTATTTGCTGGTCGTTATCGCATTCTGAAATCTATTGGCCGCGGTGGTATGG  
CGGATGTTTATTTAGCAAATGACTTGATCTTGGATAATGAAGACGTTGCAATCAAGGTCTTGCGTACCAA  
TTATCAAACAGATCAGGTAGCAGTTGCGCGTTTCCAACGAGAAGCGCGGCCATGGCTGAATTGAACCAT  
CCCAATATTGTTGCCATCCGGGATATAGGTGAAGAAGACGGACAGCAATTTT TAGTAATGGAATATGTGG  
ATGGTGCTGACCTAAAGAGATACATTCAAATCATGCTCCATTATCTAATAATGAA GTGGTTAGAATTAT  
GGAAGAAGTCCTTTCTGCTATGACTTTAGCCCAACAAAAGGAATTGTACACAGAGATTTAAAACCTCAA  
AATATCCTACTAATAAGGAGGGTGTGTCAAAGTAACTGATTTCCGCATCGCAGTAGCCTTTCAGAAA  
CAAGCTTGACACAACTAATTCGATGTTAGGCAGTGTTTACTTGTCTCCAGAACAGGCTCGCGGCTC  
CAAAGCGACGATTCAAAGTGATATTTATGCGATGGGGATTATGCTCTTTGAGATGTTGACAGGCCATATC  
CCTTATGACGGCGATAGTGCTGTTACGATTGCCTTGCAACATTTTCAAAGCCTCTTCCATCTATTATCG  
AGGAGAACCACAATGTGCCACAAGCTTTGGAGAATGTTGTTATTCGAGCAACAGCCAAGAAATTAAGTGA  
TCGTTACGGGTCAACCTTTGAAATGAGTCGTGACTTAATGACGGCGCTTAGTTATAATCGTAGTCGGGAG  
CGTAAGATTATCTTTGAGAATGTTGAAAGTACCAAACCCCTCCCCAAAGTGGCCTCAGGTCCCACCGCTT  
CTGTAAAATTGTCTCCCCCTACCCCAACAGTGTTAACACAGGAAAGTCGATTAGATCAA ACTAATCAAAC  
AGATGCTTTACAGCCCCCACCACAAAAGAAAAAAGTGGTCGTTTTTTAGGTACTTTATTCAA AATTCTT  
TTTTCTTTCTTTATTGTAGGTGTAGCACTCTTTACTTATCTTATACTAACTAAACCAACTTCTGTGAAAG  
TTCCTAATGTAGCAGGCACTAGTCTTAAAGTTGCCAAACAAGAACTGTATGATGTTGGGCTAAAAGTGGG  
TAAATCAGGCAAATTGAGAGTGATACGGTTGCTGAGGGAAATGTAGTTAGAACAGATCCTAAAGCAGGA  
ACAGCTAAGAGGCAAGGCTCAAGCATTACGCTTTATGTGTCAATTGGAAACAAAGGTTTTGACATGGAAA



ACTACAAAGGACTAGATTATCAAGAAGCTATGAATAGTTTGATAGAACTTATGGTGTTCACAAATCAAA  
AATCAAAATTGAGCGCATTGTAACTAATGAATATCCTGAAAATACAGTCATCAGTCAATCGCCAAGTGCG  
GGTGATAAATTTAATCCAAACGGAAAGTCTAAAATTACGCTCAGTGTGCTGTTAGTGATACGATCACTA  
TGCCTATGGTAACAGAATATAGTTATGCAGATGCAGTCAATACCTTAACAGCTTTAGGTATAGATGCATC  
5 TAGAATAAAAGCTTATGTGCCAAGCTCTAGCTCAGCAACGGGCTTTGTGCCAATTCAATCTCCTAGTTCT  
AAAGCTATTGTGAGTGGTCAATCTCCTTACTATGGAACGTCTTTGAGTCTGTCTGATAAAGGAGAGATTA  
GTCTTTACCTTTATCCAGAAGAAACACACTCTTCTAGTAGCTCATCGAGTTCAACGTCAAGTTCAAACAG  
TTCTTCAATAAATGATAGTACTGCACCAGGTAGCAACACTGAATTAAGCCCATCAGAACTACTTCTCAA  
ACACCTTAA

10

Preferred GAS 372 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 25; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 25, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250 or more). These GAS 372 proteins include variants (e.g. allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 25. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 25. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 25. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

15

20

#### (14) GAS 039

GAS 039 corresponds to M1 GenBank accession numbers GI:13621542 and GI:15674446, to M3 GenBank accession number GI: 21909730, to M18 GenBank accession number GI: 19745398 and is also referred to as 'Spy0266' (M1), 'SpyM3\_0194' (M3), and 'SpyM18\_0250' (M18). Amino acid and polynucleotide sequences of GAS 039 of an M1 strain are set forth below:

25

#### SEQ ID NO: 27

MDLILFLLVLVLLGLGAYLLFKVNGLOHQLAQTLEGNADNLSQMTYQLDTANKQQLLELTQLMNRQQAG  
LYQQLTDIRDVLHRSLSRDRSDKRLEKINQQVNQSLKNMQESNEKRLEKMRQIVEEKLEBTLKNRLHA  
30 SFDSVSKQLESVNKGLGEMRSVAQDVGTNLKVLSTKTRGILGELQLGQI IEDIMTSSQYEREFVTVSGS  
SERVEYAIKLPNGNGGGYIYLPIDSKFPLEDYRLEDAYEVDKLAIEASRKALLAAIKRFAKDIHKKYL  
NPPEITNFGVMFLPTEGLYSEVVRNASFFDSLRRREENIVVAGPSTLSALLNSLSVGFKTLNIQKNADDIS  
KILGNVKLEFDKFGLLAKAQKQMNANTLDQLISTRNNAIVRALNTVETQDQATKSLNMPLEEN  
35 NEN

#### SEQ ID NO: 28

ATGGACCTTATCTTGTTCCTTTTGGTCTTGGTTCTCTTAGGTTTAGGGGCTTATCTGTTGTTCAAAGTCA  
ACGGCCTTCAACATCAGCTTGCCCAAACCTAGAAAGGCAACGCGGATAATTTGTCTGACCAAATGACCTA  
40 CCAGTTGGATACAGCTAACAAACAACAAATTGTTAGAGCTAACACAGCTGATGAACCGACAACAAGCAGGC  
CTTTACCAACAATTAACAGATATTCGTGACGTCTTGACCGTAGTTTGTCTGATAGTAGGGACCGGTCTG  
ACAAACGCTTAGAAAAAATTAACCAGCAGGTCAACCAATCGCTCAAAAATATGCAAGAATCTAACGAAAA  
ACGTTTGGAGAAAATGCGCCAGATCGTTGAAGAAAAATTGGAAGAAACCTTAAAAAATCGTCTGCACGCC  
TCTTTCGATTCTGTATCCAAGCAACTAGAAAGTGTCAATAAAGGCTTGGGAGAAATGCGTAGCGTGGCTC  
45 AAGATGTGGGTACTTTAAATAAGGTTTGTCCAATACCAAAACACGAGGCATTTTAGGCGAACTTCAACT  
AGGCCAAATCATTGAGGATATCATGACATCAAGCCAGTACGAAAGAGAATTTGTAACGGTTAGTGTTCT  
AGTGAACGCGTAGAATATGCGATTAAAGCTCCAGGAAATGGTCAAGGCGGTTATATTTACCTACCGATTG  
ACTCAAAATTCCTCTTGAAGATTATTACCGATTAGAAGATGCTTACGAAGTTGGTGATAAACTGGCCAT  
CGAGGCTAGCCGAAAAGCACTTCTGGCAGCTATCAAACGCTTTGCCAAAGACATTCATAAAAAGTACTTG  
50 AACCCCCCAGAGACGACCAATTTCCGAGTTATGTTCTTACCAACAGAAGGTCTTTATTCAGAAGTGGTCA  
GAAATGCGTCTTTCTTGATAGCCTTCGTGCGGAAGAAAATATTGTGGTTGCAGGCCCTTCGACCCTGTC



TGCTTTGCTGAATTCCTTATCTGTTGGTTTCAAGACCCTTAATATCCAAAAAATGCTGATGACATCAGT  
AAAATTTTAGGCAATGTCAAGTTAGAATTCGATAAATTTGGCGGCCTGCTTGCCAAGGCTCAAAAACAAA  
TGAATACAGCTAATAATACGCTGGATCAGCTCATTTCACCAAGGACAAATGCCATTGTTTCGAGCCTTGAA  
TACCGTTGAAACTTATCAAGACCAAGCAACAAAATCTCTCTTGAACATGCCCTTATTAGAAGAGGAAAAT  
AATGAAAATTAA

Preferred GAS 039 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 27; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 27, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, or more). These GAS 039 proteins include variants (e.g. allelic variants, homologs, orthologs, paralog, mutants, etc.) of SEQ ID NO: 27. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 27. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 27. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

#### (15) GAS 042

GAS 042 corresponds to M1 GenBank accession numbers GI:13621559 and GI:15674461, to M3 GenBank accession number GI: 21909745, to M18 GenBank accession number GI: 19745415, and is also referred to as 'Spy0287' (M1), 'SpyM3\_0209' (M3), and 'SpyM18\_0275' (M18). Amino acid and polynucleotide sequences of GAS 042 of an M1 strain are set forth below:

#### SEQ ID NO: 29

MTKEKLVAFSQAHABPAWLQERRLAALAEAI PNLELPTIERVKFHRWN LGDGLTENESLASVPDFIAIGD  
NPKLVQVGTQT VLEQLPMALIDKGVVFSDFYTALEEIPEVIEAHFGQALAFDEDKLAAYHTAYFN SAAVL  
YVPDHLEITTP IEAIFLQSDSDVPFNKHLVVIAGKESKFTYLERFESIGNATQKISANISVEVIAQAGS  
QIKFSAIDRLGPSVTYIISRRGRLEKDANIDWALAVMNEGNVIADFDSDLIGQGSQADLKVVAASSGRQV  
QGIDTRVTNYGQRTVGHI LQHGVILERGTLT FNIGI GHILKDAKGADAQQESRVLMLSDQARADANPILLI  
DENEVTAGHAASIGQVDPEDMYILMSRGLDQETAERLVIRGFLGAVIAEIP IPSVRQBI IKVLDEKLLNR

#### SEQ ID NO: 30

ATGACAAAAGAAAACTAGTGGCTTTTTTCGCAAGCCCACGCTGAGCCTGCTTGGCTGCAAGAACGGCGTT  
TAGCGGCATTAGAAGCCATTCCAAATTTGGAATTACCAACCATCGAAAGGGTTAAATTTACCGTTGGAA  
TCTAGGAGATGGTACCTTAACAGAAAATGAAAGTCTAGCTAGTGTTCCAGATTTTATAGCTATTGGAGAT  
AACCCAAAGCTTGTT CAGGTAGGCACGCAACAGTCTTAGAACAGTTACCAATGGCGTTAATTGACAAGG  
GAGTTGTTTT CAGTGATTTTTATACGGCGCTTGAGGAAATCCCAGAAGTAATTGAAGCTCATTTTGGTCA  
GGCATTAGCTTTTGATGAAGACAACTAGCTGCCTACCACACTGCTTATTTTAATAGCGCAGCCGTGCTC  
TACGTTCTTGATCACTTGGAATCACAACCTCTATTGAAGCTATTTTCTTACAAGATAGTGACAGTGACG  
TTCCTTTTAACAAGCATGTTCTAGTGATTGCAGGAAAAGAAAGTAAGTTCACCTATTTAGAGCGTTTTGA  
ATCTATTGGCAATGCCACTCAAAGATCAGCGCTAATATCAGTGTAGAAGTGATTGCTCAAGCAGGCAGC  
CAGATTAAATTCTCGGCTATCGACCGCTTAGGTCTTCAGTGACAACCTATATTAGCCGTGAGGACGTT  
TAGAGAAGGATGCCAACATTGATTGGGCCTTAGCTGTGATGAATGAAGGCAATGTCATTGCTGATTTTGA  
CAGTGATTTGATTGGTCAGGGCTCACAAGCTGATTGAAAGTTGTTGCAGCCTCAAGTGGTCGTCAGGTA  
CAAGGTATTGACACGCGCGTGACCAACTATGGTCAACGTACGGTCGGTCATATTTTACAGCATGGTGTGA  
TTTTGGAACGTGGCACCTTAACGTTTAACGGGATTGGTCATATTCTAAAAGACGCTAAGGGAGCTGATGC  
TCAACAAGAAAGCCGTGTTTTGATGCTTTCTGACCAAGCAAGAGCCGATGCCAATCCAATCCTCTTAATT  
GATGAAAATGAAGTAACAGCAGGT CATGCAGCTTCTATCGGT CAGGTTGACCCTGAAGATATGTATTACT  
TGATGAGTCGAGGACTGGATCAAGAAACAGCAGAACGATTGGTTATTAGAGGATTCCTAGGAGCGGTTAT  
CGCTGAAATTCCTATTCCATCAGTCCGCCAAGAGATTATTAAGGTTTTAGATGAGAAATTGCTTAATCGT  
TAA



Preferred GAS 042 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 29; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 29, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, or more). These GAS 042 proteins include variants (e.g. allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 29. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 29. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 29. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

**(16) GAS 058**

GAS 058 corresponds to M1 GenBank accession numbers GI:13621663 and GI:15674556, to M3 GenBank accession number GI: 21909841, to M18 GenBank accession number GI: 19745567 and is also referred to as 'Spy0430' (M1), 'SpyM3\_0305' (M3), and 'SpyM18\_0477' (M18). Amino acid and polynucleotide sequences of GAS 058 of an M1 strain are set forth below:

**SEQ ID NO: 31**

MKWSGFMKTKSKRFLNLATLCLALLGTTLLMAHPVQAEVISKRDYMTFRGLGLDLEDDSANYPNLEARYK  
GYLEGYBKGLKGDDIPERPKIQVPEDVQPSDHGDYRDGYEEGFGEQHKRDPLETEABDDSQGGRQEGRQ  
GHQEGADSSDLNVEESDGLSVIDEVVGVIYQAFSTIWTYLSGLF

**SEQ ID NO: 32**

ATGAAATGGAGTGGTTTTATGAAAACAAAATCAAAACGCTTTTAAACCTAGCAACCCTTTGCTTGGCCC  
TACTAGGAACAACCTTTGCTAATGGCACATCCCGTACAGGCGGAGGTGATATCAAAAAGAGACTATATGAC  
TCGCTTCGGGTTAGGCGATTAGAAAGATGATTGAGCTAACTATCCTTCAAATTTAGAAGCTAGATATAAA  
GGATATTTAGAGGGATATGAAAAGGCTTAAAAGGAGATGATATACCCGAACGGCCCAAGATTCAGGTTC  
CTGAGGATGTTGAGCATCTGACCATGGCGACTATAGAGATGGTTATGAGGAAGGATTTGGAGAAGGACA  
ACATAAACGTGATCCATTAGAAACAGAAGCAGAAGATGATTCTCAAGGAGGACGTCAAGAAGGACGTCAA  
GGACATCAAGAAGGAGCAGATTCTAGTGATTGTAACGTTGAAGAAAGCGACGGTTTGTCTGTTATTGATG  
AAGTAGTTGGAGTAATTTATCAAGCATTTAGTACTATTTGGACATACTTAAGCGGTTTGTCTCTAA

Preferred GAS 058 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 31; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 31, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, or more). These GAS 058 proteins include variants (e.g. allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 31. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 31. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 31. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 31 is removed. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).



**(17) GAS 290**

GAS 290 corresponds to M1 GenBank accession numbers GI:13622978 and GI:15675757, to M3 GenBank accession number GI: 21911221, to M18 GenBank accession number GI: 19746869 and is also referred to as 'Spy1959' (M1), 'SpyM3\_1685' (M3), and 'SpyM18\_2026' (M18). Amino acid and polynucleotide sequences of GAS 290 of an M1 strain are set forth below:

**SEQ ID NO: 33**

MKHILFIVGSLREGSFNHQLAAQAQKALEHQAVVSYLNWKDVPVLNQDIEANAPLPVVDARQAVQSADAI  
WIFTPVYNFSIPGSVKNLLDWLSRALDLSIPTGSAIGGKVVTVSSVANGGHDQVFDQFKALLPFIRTSV  
AGEFTKATVNPDAWGTGRLBISKETKANLLSQAEALLAAI

**SEQ ID NO: 34**

ATGAAACATATTTTATTTATTTGTTGGCTCGCTTCGTGAAGGGTCTTTTAACCATCAATTAGCGGCTCAAG  
CACAAAAGCTCTGGAACATCAAGCAGTTGTATCTTACTTAAATTGGAAAGACGTTCTGTGTTTGAATCA  
AGATATCGAAGCTAATGCACCTTTACCAGTTGTTGACGCTCGTCAAGCTGTTTCAAGTCAAGCGGATGCTATC  
TGGATTTTACACCAGTTTACAACCTTCTCTATTCCAGGTTCTGTTAAAAACCTGCTAGACTGGTTGTCTC  
GTGCTCTTGATTTGTCTGATCCGACGGGCCCCTCTGCTATTGGCGGTAAGGTGGTTACGGTCTCTTCAGT  
TGCAAATGGCGGGCATGATCAAGTATTTGATCAGTTTAAAGCACTATTGCCGTTTATCCGAACTTCAGTA  
GCAGGAGAGTTTACAAAAGCAACTGTGAATCCTGATGCCTGGGGAACAGGAAGGCTTGAGATTTCAAAG  
AGACAAAAGCAAACCTTGCTATCTCAGGCAGAGGCTCTTTTAGCGGCTATTTAG

Preferred GAS 290 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 33; and/or (b) which is a fragment of at least *n* consecutive amino acids of SEQ ID NO: 33, wherein *n* is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 or more). These GAS 290 proteins include variants (e.g. allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 33. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 33. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 33. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

**(18) GAS 511**

GAS 511 corresponds to M1 GenBank accession numbers GI:13622798 and GI:15675592, to M3 GenBank accession number GI: 21911053, to M18 GenBank accession number GI: 19746700 and is also referred to as 'Spy1743' (M1), 'SpyM3\_1517' (M3), 'SpyM18\_1815' (M18) and 'accA'. Amino acid and polynucleotide sequences of GAS 511 of an M1 strain are set forth below:

**SEQ ID NO: 35**

MTDVSRIKLEARDQGRLLTLDYANLIFDDFMELHGDHFSDDGAIVGGLAYLAGQPVTVIGIQKGKNLQD  
NLARNFGQPNPEGYRKALRLMKQAEKFRPVVTFINTAGAYPGVGAEBRGQGEAIAKNLMEMSDLKVPII  
AIIIGEGSGGALALAVADQVWMLNTMYAVLSPEGFASILWKDGSRAEABLMKITAGELYKMGIVDR  
IIPHEGYFSSSEIVDIIKANLIEQITSLQAKPLDQLLDERYQFRKY

**SEQ ID NO: 36**

ATGACAGATGTATCAAGAATTTTAAAAGAAGCGCGTGATCAAGGGCGTTTAAACAACCTTTGGATTACGCCA  
ACCTTATTTTCGATGACTTTATGGAACGTCATGGCGATCGCCATTTTTCAGATGATGGTGCCATTGTAGG  
TGGCCTAGCTTATTTGGCGGGACAACCTGTTACGGTCATTGGTATTCAAAAAGGTAAGAATTTACAGGAT  
AATTTGGCAAGGAATTTTGGCCAGCCCAATCCAGAAGGTTATCGTAAAGCTTTGCGCCTTATGAAACAGG







AGCATCACGTGGTATGGGAACGCGTTTGGAGTTACGTTTCGGTTGATCCGACAGCTAATCCTTATTTAGCC  
 TTGGCTGTTCTCTTGGGAAGCTGGATTAGATGGTATCATTAAACAAAATTGAAGCTCCAGAACCCGTTGAAG  
 CTAACATTTATACCATGACAATGGAAGAACGAAATGAAGCAGGCATTATTGATTGTCATCAACGCTTCA  
 TAATGCCTTAAAAGCTCTTCAAAAAGATGATGTGGTACAAAAGGCACTAGGTTACCATATCTACACTAAT  
 5 TTCTTAGAAGCAAAACGAATTGAATGGTCTTCCTATGCAACTTTTGTCTCAATGGGAAATTGACCATT  
 ATATTCATAATTATTAG

Preferred GAS 533 proteins for use with the invention comprise an amino acid sequence: (a) having  
 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
 10 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 37; and/or (b) which is a fragment of at least  $n$   
 consecutive amino acids of SEQ ID NO: 37, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 533 proteins include variants (e.g.  
 allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 37. Preferred fragments  
 of (b) comprise an epitope from SEQ ID NO: 37. Other preferred fragments lack one or more amino  
 15 acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more  
 amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID  
 NO: 37. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide,  
 of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

#### (20) GAS 527

20 GAS 527 corresponds to M1 GenBank accession numbers GI:13622332, GI:15675169, and  
 GI:24211764, to M3 GenBank accession number GI: 21910381, to M18 GenBank accession number  
 GI: 19746136, and is also referred to as 'Spy1204' (M1), 'SpyM3\_0845' (M3), 'SpyM18\_1155'  
 (M18) and 'guaA'. GAS 527 has also been identified as a putative GMP synthetase (glutamate  
 hydrolyzing) (glutamate amidotransferase). Amino acid and polynucleotide sequences of GAS 527 of  
 25 an M1 strain are set forth below:

#### SEQ ID NO: 39

MTEISILNDVQKIIIVLDYGSQYNQLIARRIREFGVFSELKSHKITAQELREINPIGIVLSGGPNSVYADN  
 AFGIDPEIFELGIPILGICYGMQLITHKLGGKVVPAGQAGNREYGQSTLHLRETSKLFSGTPQEQLVLM  
 HGDVTEIPEGFHLVGDSNDCPYAAIENTEKNLYGIQFHPVRHSVYGNLILKNFAISICGARGDWSMDN  
 30 FIDMEIAKIRETVGDRKVLGLSGGVDSSVVGVLQKAIGDQLTCIFVDHGLLRKDEGDQVMGMLGGKFG  
 LNIIRVDASKRFLDLLADVEDPEKKRKIIGNEFVYVFDDEASKLKGVDFLAQGTLYTDIIESGTETAQTI  
 KSHHNVGGLPEDMQFELIPLNTLFKDEVRLGIALGMPEEIVWRQFPFPGPLAIRVMGAITEEKLETVR  
 ESDAILRBEIAKAGLDRDVWQYFTVNTGVRVSGVMGDGRTYDYTIAIRAITSIDGMTADFAQLPVDVLKK  
 35 ISTRIVNEVDHVNRIYDITSKPPATVEWE

#### SEQ ID NO: 40

ATGACTGAAATTTCAATTTTGAATGATGTTCAAAAAATTATCGTTCTTGATTATGGTAGCCAGTACAATC  
 AGCTTATTGCTAGACGATTTCGAGAGTTTGGTGTCTTCTCCGAACTAAAAAGCCATAAAATCACCGCTCA  
 AGAACTTCGTGAGATCAATCCCATAGGTATCGTTTTATCAGGAGGGCCTAACTCTGTTTACGCTGATAAC  
 40 GCCTTTGGCATTGACCCTGAAATCTTGAAGTAGGGATTCCGATTCTTGGTATCTGTTACGGTATGCAAT  
 TAATCACCCATAAATTAGGTGGTAAAGTTGTTCTGCTGGACAAGCTGGTAATCGTGAATACGGTCAGTC  
 AACCTTCATCTTCGTGAAACGTCAAAATTATTTTCAGGCACACCTCAAGAACAACCTCGTTTTGATGAGC  
 CATGGTGATGCTGTTACTGAAATCCAGAAGTTTCCACCTTGTGGAGACTCAAATGACTGTCCCTATG  
 CAGCTATTGAAAATACTGAGAAAACTTTACGGTATTCAGTTCACCCAGAAGTGAGACACTCTGTTTA  
 45 TGGAATGACATTCTTAAAACTTTGCTATATCAATTTGTGGCGCGGTGGTGAATGGTCAATGGATAAT  
 TTTATTGACATGGAAATTGCTAAATTCGTGAAACTGTAGGCGATCGTAAAGTTCTTCTAGGTCTTTCTG  
 GTGGAGTTGATTCTTCAAGTTGTTGGTGTCTACTTCAAAAAGCTATCGGTGACCAATTAATTGTATTTT  
 CGTTGATCACGGTCTTCTTCGTAAAGACGAGGGCGATCAAGTTATGGGAATGCTTGGGGGCAAATTTGGC  
 CTAATATTATCCGTGTGGATGCTTCAAAACGTTTCTTAGACCTTCTTGCAGACGTTGAAGATCCTGAGA



AAAAACGTA AAAATTATTGGTAATGAATTTGTCTATGTTTTTGATGATGAAGCCAGCAAATTAAAAGGTGT  
TGACTTCCTTGCCCAAGGAACACTTTATACTGATATCATTGAGTCAGGAACAGAACTGCTCAAACCATC  
AAATCACATCACAAATGTGGGTGGTCTCCCCGAAGACATGCAGTTTGAATTGATTGAGCCCTTAAACACTC  
TTTTCAAAGATGAAGTTCGAGCGCTTGGAAATCGCTCTTGGAAATGCCTGAAGAAATTGTTTGGCGCCAACC  
ATTTCCAGGTCTTGGACTTGCTATCCGTGT CATGGGAGCAATTACTGAAGAAAACTTGAAACCGTTTCGC  
GAATCAGACGCTATCCTTCGTGAAGAAATTGCTAAGGCTGGACTTGATCGTGACGTGTGGCAATACTTTA  
CAGTTAACACAGGTGTCCGTTCTGTAGGCGTCATGGGAGATGOTCGTACTTATGATTATACCATCGCCAT  
TCGTGCTATTACGTCTATTGATGGTATGACAGCTGACTTTGCTCAACTTCCTTGGGATGTCTTGAAAAA  
ATCTCAACACGTATCGTAAATGAAGTTGACCACGTTAACCGTATCGTCTACGACATCACAAGTAAACCAC  
CCGCAACAGTTGAATGGGAATAA

**(21) GAS 294**

**SEQ ID NO: 41**

**SEQ ID NO: 42**

-21-



AATGCAGCTGGAAGCCTTTATAATATCGTTGGTTTCCAAACCCATCTCAAATGGGGTGAGCAAAAACGCG  
TTTTCCAAATGATTCCAGGGCTTGAAAATGCTGAGTTTGTCCGCTACGGCGTCATGCATCGCAATTCCTA  
TATGGATTACCAAATCTTTTAACCGAAACCTTCCAATCTCGGAGCAATCCAAACCTTTTCTTTGCAGGT  
CAGATGACTGGAGTTGAAGGTTATGTCGAATCAGCTGCTTCAGGTTTAGTAGCAGGAATCAATGCTGCTC  
5 GTTTGTTCAAAAAGAGAAGAAGCACTTATTTTTCCTCAGACAACAGCTATTGGGAGTTTGCCTCATTATGT  
GACTCATGCCGACAGTAAGCATTTCACCAATGAACGTCAACTTTGGCATCATCAAAGAGTTAGAAGGC  
CCACGCATTCTGTGACAAAAAGAAGCTTATGAAGCTATTGCTAGTCGTGCTTTGGCAGATTTAGACACCT  
GCTTAGCGTCGCTTTAA

- 10 Preferred GAS 294 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 41; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 41, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 294 proteins include variants (e.g.
- 15 allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 41. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 41. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 41. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide,
- 20 of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

## (22) GAS 253

GAS 253 corresponds to M1 GenBank accession numbers GI:13622611, GI:15675423, and GI:21362716, to M3 GenBank accession number GI: 21910711, to M18 GenBank accession number GI: 19746473 and is also referred to as 'Spy1524' (M1), 'SpyM3\_1175' (M3), 'SpyM18\_1541' (M18) and 'murG'. GAS 253 has also been identified as a putative undecaprenyl-PP-MurNAc-pentapeptide-UDPGlcNAc GlcNAc transferase. Amino acid and polynucleotide sequences of GAS 253 of an M1 strain are set forth below:

### SEQ ID NO: 43

MPKKILFTGGGTVGHVTLNLILIPKFIKDGWEVHYIGDKNGIEHTEIEKSGLDVTFHAIATGKLRRYFSW  
30 QNLADVFKVALGLLQSLFIVAKLRPQALFSKGGFVSVPPVVAALLGKPVFIHESDRSMGLANKIAYKFA  
TTMYTTFEQBDQLSKVKHLGAVTKVFKDANQMPBESTQLEAVKEYFSRDLKTLPLIGGSAGAHVFNQFISD  
HPBLKQRYNIINITGDPHLNELSSHLRYVDYVTDLYQPLMAMADLVVTRGGSNTLPELLAMAKLHLIVPL  
GKEASRGDQLENATYFEKRGYAKQLQEPDLTLHNFQAMADLFEHQADYEATMLATKEIQSPDFFYDLLR  
ADISSAIKEK

### SEQ ID NO: 44

ATGCCTAAGAAGATTTTATTACAGGTGGTGGAACTGTAGGTCATGTCACCTTGAACCTCATTCTCATAC  
CAAAATTTATCAAGGACGGTTGGGAAGTACATTATATTGGTGATAAAAATGGCATTGAACATACAGAAAT  
TGAAAAGTCAGGCCTTGACGTGACCTTTCATGCTATCGCGACAGGCAAGCTTAGACGCTATTTTTCATGG  
40 CAAAATCTAGCTGATGTTTAAAGGTTGCACTTGGCCTCCTACAGTCTCTCTTTATTGTTGCCAAGCTTC  
GCCCTCAAGCCCTTTTTTCCAAAGGTGGTTTGTCTCAGTACCGCCAGTTGTGGCTGCTAAATTGCTTGG  
TAAACCAGTCTTTATTCATGAATCAGATCGGTCAATGGGACTAGCAAACAAGATTGCCTACAAATTTGCA  
ACTACCATGTATACCACTTTTGAGCAGGAAGACCAGTTGTCTAAAGTTAAACACCTTGAGCGGTGACAA  
AGGTTTTCAAGATGCCAACCAAAATGCCTGAATCAACTCAGTTAGAGGCGGTGAAAGAGTATTTTAGTAG  
45 AGACCTAAAAACCCTCTGTTTATTGGTGGTTCGGCAGGGGCGCATGTGTTAATCAGTTTATTAGTGAT  
CATCCAGAATTGAAGCAACGTTATAATATCATCAATATTACAGGAGACCCTCACCTTAATGAATTGAGTT  
CTCATCTGTATCGAGTAGATTATGTTACCGATCTCTACCAACCTTTGATGGCGATGGCTGACCTTGTAGT  
GACAAGAGGGGGCTCTAATACTTTTGTAGCTACTGGCAATGGCTAAGCTACACCTCATCGTTCCTCTT  
GGTAAAGAAGCTAGCCGTGGCGATCAGTTAGAAAATGCCACTTATTTTGAGAAGAGGGGCTACGCTAAAC



AATTACAGGAACCTGATTAACTTTGCATAATTTTGATCAGGCAATGGCTGATTGTTTGAACATCAGGC  
TGATTATGAGGCTACTATGTTGGCAACTAAGGAGATTCAGTCACCGGACTTCTTTTATGACCTTTTGAGA  
GCTGATATTAGCTCCGCGATTAAGGAGAAGTAA

- 5 Preferred GAS 253 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 43; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 43, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 253 proteins include variants (e.g. 10 allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 43. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 43. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 43. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, 15 of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

### (23) GAS 529

- GAS 529 corresponds to M1 GenBank accession numbers GI:13622403, GI:15675233, and GI:21759132, to M3 GenBank accession number GI: 21910446, to M18 GenBank accession number GI: 19746203 and is also referred to as 'Spy1280' (M1), 'SpyM3\_0910' (M3), 'SpyM18\_1228' 20 (M18) and 'glmS'. GAS 529 has also been identified as a putative L-glutamine-D-fructose-6-phosphate aminotransferase (Glucosamine-6-phosphate synthase). Amino acid and polynucleotide sequences of GAS 529 of an M1 strain are set forth below:

#### SEQ ID NO: 45

- 25 MCGIVGVGNRNATDILMQGLEKLEYRGYDSAGIFVANANQTNLIKSVGRIADLRKIGIDVAGSTGIGH  
TRWATHGQSTEDNAHPHTSQTGRFVLVHNGVIENYLHIKTEFLAGHDFKGQTDTEIAVHLIGKFVEBDKL  
SVLEAFKKSLSIIEGSYAFALMDSQATDTIYVAKNKSPLLI GLGEGYNMVCSDAMAMIRETSEFMEIHDK  
ELVILTKDKVTVDYDGKELIRDSYTAELDLSDIGKGYPPFYM LKEIDEQPTVMRQLISTYADETGNVQV  
DPAIITSIQEADRLYILAAGTSYHAGFATKNMLEQLTDPVBLGVASEWGYHMP LLSKKPMFILLSSQSGE  
TADSRQVLVKANAMGIPSLTVTNVPGSTLSREATYTMLIHAGPEIAVASTKAYTAQIAALAF LAKAVGEA  
30 NGKQEALDFNLVHEL SLVAQSI EATLSEKDLVAEKVQALLATTRNAFYIGRNDYYVAMEAALKLKEISY  
IQCEGFAAGELKHGTISLIEEDTPVIALISSQLVASHTRGNIQEVAARGAHVLT VVEEGLDREGDDIIV  
NKVHPFLAPIAMVIPTQLIAYYASLQRLDVKPRNLAKAVTVE

#### SEQ ID NO: 46

- 35 ATGTGTGGAATTGTTGGAGTTGTTGGAAATCGCAATGCAACGGATATTTTAATGCAAGGCCTTGAAAAGC  
TTGAATACCGGGGTATGATTGAGCAGGAATTTTGTGGCTAATGCCAATCAAACAACTTGATTAAATC  
AGTGGGGCGGATTGCTGATTGCGTGCCAAGATTGGCATTGATGTTGCTGGTTCAACAGGGATTGGTCAC  
ACCCGTTGGGCAACGCATGGCCAATCAACAGAGGATAATGCCCATCCTCACACGTCACAACTGGACGTT  
TTGTACTTGTTTATAATGGTGTGATTGAAAATTACCTTACATTAAACAGAGTTCCTAGCTGGACATGA  
40 TTTTAAGGGGCAGACAGATACTGAGATTGCAGTACACTTGATTGGAAAATTTGTGGAAGAAGACAAGTTG  
TCAGTACTGGAAGCTTTTAAAAAATCTTTAAGCATTATTGAAGGTTCTACGCCTTTGCATTAAATGGATA  
GCCAAGCAACTGATACTATTTATGTGGCTAAAAACAAGTCTCCATTGTTGATTGGACTTGGTGAAGGTTA  
CAACATGGTTTGTTCAGATGCCATGGCCATGATTCGTGAAACCAGTGAATTTATGGAAATTCATGATAAG  
GAGCTAGTTATTTTAACCAAAGATAAGGTAAGTGTACAGACTACGATGGTAAAGAGCTGATACGAGATT  
45 CCTACACTGCTGAATTAGACTTATCTGATATTGGCAAAGGGACTTATCCTTTCTATATGCTGAAAGAAAT  
TGATGAGCAACCAACCGTAATGCGTCAATTAATTTCAACTTATGCAGATGAACTGGTAACGTACAGGTT  
GATCCGGCTATCATTACCTCTATCCAAGAGGCTGACCGTCTTTATATTTTAGCGGCAGGGACTTCCTACC  
ATGCTGGTTTTGCAACAAAAAATATGCTTGAGCAATTGACAGATACACCAGTTGAGTTGGGCGTGGCTTC  
TGAGTGGGGTTACCACATGCCTCTGCTTAGCAAGAAACCAATGTTTATTCTACTAAGCCAATCAGGAGAA



ACCGCAGATAGTCGTCAAGTTTTAGTAAAGGCAAATGCTATGGGCATTCCGAGTTTGACAGTAACTAACG  
 TTCCAGGATCAACCTTATCACGTGAAGCAACATACACCATGTTGATTATGCTGGACCTGAAATTGCTGT  
 TGGCTCTACAAAAGCTTACACTGCACAAATTGCTGCCCTTGCCTTTTTGGCTAAGGCAGTTGGTGAGGCA  
 AATGGTAAGCAAGAAGCTCTTGACTTTAACTTGGTACATGAGTTGTCATTGGTTGCCCAATCTATTGAGG  
 5 CGACTTTGTCTGAAAAAGATCTCGTGGCAGAAAAGGTTCAAGCTTTGCTAGCTACTACTCGTAATGCTTT  
 TTACATCGGGCGTGGCAATGATTATTACGTTGCGATGGAAGCTGCTTTGAAATTAAGAGATTCTTAT  
 ATTCAATGCGAAGGCTTTGCGGCTGGTGAATTGAAACATGGAACCATTTCAATTAATTGAGGAGGACACGC  
 CAGTAATCGCTTTAATATCGTCTAGTCAGTTGGTTGCCTCTCATACGCGTGGTAATATTCAAGAAGTTGC  
 TGGCCGTGGGGCTCATGTTTTAACAGTTGTGGAAGAAGGGCTTGACCGTGAGGGAGATGACATTATTGTC  
 10 AATAAGGTTTCATCCTTTCCTAGCCCCGATTGCTATGGTCATTCCAACCTCAACTGATTGCTTACTACGCTT  
 CATTACAACGTGGACTTGATGTTGATAAGCCACGTAATTTGGCTAAAGCTGTAACAGTAGAATAA

Preferred GAS 529 proteins for use with the invention comprise an amino acid sequence: (a) having  
 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
 15 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 45; and/or (b) which is a fragment of at least  $n$   
 consecutive amino acids of SEQ ID NO: 45, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 529 proteins include variants (e.g.  
 allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 45. Preferred fragments  
 of (b) comprise an epitope from SEQ ID NO: 45. Other preferred fragments lack one or more amino  
 20 acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more  
 amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID  
 NO: 45. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide,  
 of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

#### (24) GAS 045

25 GAS 117 corresponds to M3 GenBank accession number GI: 21909751, M18 GenBank accession  
 number GI: 19745421 and is referred to as 'SpyM3\_0215' (M3), 'SpyM18\_oppA' (M18) and 'oppA'.  
 GAS 045 has been identified as an oligopeptide permease. Amino acid and polynucleotide sequences  
 of GAS 045 from an M1 strain are set forth below:

#### SEQ ID NO: 47

30 VTFMKSKWLAASVAILSVSALAACGNKNASGGSEATKTYKYVFNPKSLDYILTNGG  
 GTTDVITQMVDGLLENDEYGNLVPVSLAKDWKSKDGLTYTYTLRDGVSWYTADGEBYAPV  
 TAEDFVTGLKHAVDDKSDALYVVEDSIKNLKAYQNGEVDPFKEVGVKALDDKTQYTLNKP  
 ESYWNSKTTYSVLPVNAKFLKSKGKDFGTTDPSSI LVNGAYFLSAFTSKSSMEFHKNEN  
 YWDAKNVGIESVKLTYS DGS DPGSFYKNFDKGEFSVARLYPNDPTYKSAKKNYADNITYG  
 35 MLTGDIRHLTNLNRTSFKNTKKDPAQQDAGKKALNNKDFRQAIQFAFDRASFQAQTAGQ  
 DAKTKALRNMLVPPTFVTIGESDFGSEVEKEMAKLGDEWKDVNLADAQDGFYNPEKAKAE  
 FAKAKEALTAEGVTFPVQLDYPVDQANAATVQEAQSFQKQSVESLGKENVIVNVLETETS  
 THEAQGFYAETPEQQDYDISSWNGPDYQDPRTYLDIMSPVGGGSVIQKLGIKAGQNKDV  
 VAAAGLDYQTLLEAAAITDDNDARYKAYAKAQAYLTDNAVDIPVVALGGTPRVTKAVP  
 40 FSGGFSWAGSKGPLAYKGMKLQDKPVTVKQYEKAKEKWMKAKAKSNAKYAEKLADHVEK

#### SEQ ID NO: 48

GTGACTTTTATGAAGAAAAGTAAATGGTTGGCAGCTGTAAGTGTGCGATCTTGTCAGTA  
 TCCGCTTTGGCAGCTTGTGGTAATAAAAATGCTTCAGGTGGCTCAGAAGCTACAAAACC  
 45 TACAAGTACGTTTTTGTAAACGATCCAAAATCATTGGATTATATTTTGACTAATGGCGGT  
 GGAACGACTGATGTGATAACACAAATGGTTGATGGTCTTTTGAAAACGATGAGTATGGT  
 AATTTAGTACCATCACTTGCTAAAGATTGGAAGGTTTCAAAGACGGTCTGACTTATACT  
 TATACTCTTCGCGATGGTGTCTCTTGGTATACGGCTGATGGTGAAGAATATGCCCCAGTA  
 ACAGCAGAAGATTTTGTGACTGGTTTGAAGCACGCGGTTGACGATAAATCAGATGCTCTT  
 50 TACGTTGTTGAAGATTCAATAAAAACTTAAAGGCTTACCAAAATGGTGAAGTAGATTTT  
 AAAGAAGTTGGTGTCAAAGCCCTTGACGATAAACTGTTCACTACTTTGAACAAGCCT







IKTYQLKGEDATLSDGGSYQIVTSNHLLEIQNRIRTELGLHKVNQLKTNATVYENLYGSTKSQTVNNNYD  
SSGQAPSYSDSHSSYANYSSGVDTCQSASTDQDSTASSHRPATPSSSSDALADESSSSSGSGSLVPPANI  
NPQT

5 **SEQ ID NO: 50**

ATGAAAATTGGAAAAAATAGTTTAAATGTTTACAGCTATTGTGTTAACAACCTGCTTGGCATTAGGTG  
TCTATCTAACTAGTGCTTATACCTTCTCAACAGGAGAAATTATCAAAGACCTTTAAAGATTTTTCGACATC  
TTCAAACAAAAGTGATGCCATTAAACAAAACAAGAGCTTTTTCTATCTTGTGATGGGTGTTGATACAGGC  
TCTTCAGAGCGTGCTCCAAGTGGGAAGGAAACAGTGATTGATGATTTTGGTTACGGTTAATCCAAAGA  
10 CCAAGAAAACAACCTATGACTAGTTTAGAACGAGATACCTTAACCACGTTATCTGGACCCAAAAATAATGA  
AATGAATGGTGTGTAAGCTAAGCTTAACGCTGCTTATGCAGCAGGTGGCGCTCAGATGGCTATTATGACC  
GTGCAAGATCTTTTGAATATCACCATTGATAACTATGTTCAAATTAATATGCAAGGCCTTATTGATCTTG  
TGAATGCAGTTGGAGGGATTACAGTTACAAATGAGTTTGATTTTCTATCTCGATTGCTGAAAACGAACC  
TGAATATCAAGCTACTGTTGCGCCTGGAACACACAAAATTAACGGTGAACAAGCTTTGGTTTATGCTCGT  
15 ATGCGTTATGATGATCCTGAGGGAGATTATGGTGCACAAAAGCGTCAACGTGAAGTCATTCAAAGGTAT  
TGAAAAAATCCTTGCTCTTGATAGCATTAGCTCTTATCGGAAGATTTTATCTGCTGTAAGTAGTAATAT  
GCAAACGAATATCGAAATCTCTCTCGCACTATCCCTAGTCTATTAGGTTATCGTGACGCACTTAGAACT  
ATTAAGACTTATCAACTAAAAGGAGAAGATGCCACTTTATCAGATGGTGGATCATACCAAATTGTTACCT  
CTAATCATTGTGTTAGAAATCCAAAATCGTATCCGAACAGAATTAGGACTTCATAAGGTTAATCAATTAAA  
20 AACAAATGCTACTGTTTATGAAAATTTGTATGGGTCAACTAAGTCTCAGACAGTAAACAACAACCTATGAC  
TCTTCAGGCCAGGCTCCATCTTATTCTGATAGTCATAGCTCTTACGCTAATTATTCAAGTGGAGTAGATA  
CCGGCCAGAGTGCTAGTACAGACCAGGACTCTACTGCTTCAAGCCATAGGCCAGCTACGCCGTCTTCTTC  
ATCAGATGCTTTAGCAGCTGATGAGTCTAGCTCATCAGGGTCTGGATCATTAGTTCCTCTGCTAATATC  
AACCCTCAGACCTAA

25 Preferred GAS 095 proteins for use with the invention comprise an amino acid sequence: (a) having  
50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 49; and/or (b) which is a fragment of at least *n*  
consecutive amino acids of SEQ ID NO: 49, wherein *n* is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
30 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 095 proteins include variants (e.g.  
allelic variants, homologs, orthologs, paralogs, mutants, *etc.*) of SEQ ID NO: 49. Preferred fragments  
of (b) comprise an epitope from SEQ ID NO: 49. Other preferred fragments lack one or more amino  
acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more  
amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID  
35 NO: 49. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of  
SEQ ID NO: 49 is removed. Other fragments omit one or more domains of the protein (e.g. omission  
of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular  
domain).

(26) **GAS 193**

40 GAS 193 corresponds to M1 GenBank accession numbers GI:13623029 and GI:15675802, to M3  
GenBank accession number GI: 21911267, to M18 GenBank accession number GI: 19746914 and is  
also referred to as 'Spy2025' (M1), 'SpyM3\_1731' (M3), 'SpyM18\_2082' (M18) and 'isp'. GAS 193  
has also been identified as an immunogenic secreted protein precursor. Amino acid and  
polynucleotide sequences of GAS 193 of an M1 strain are set forth below:

45 **SEQ ID NO: 51**

MKKRKLAVTLLSTILLNSAVPLVVADTSLRNSTSSDQPTTADTDDESETPKKDKKSKETASQHDQ  
KDHKPSHTPTPPSNDTKQTDQASSEATDKPNKDKNDTKQPDSSDQSTPSPKQSSQKESQNKDGRPTPS  
PDQQKDQTPDKTPEKSADKTPEKGPEKATDKTPEPNRDAPKPIQPPLAAAPVFI PWRESKDLSKLPSS  
RSSAAYVRHWTGDSAYTHNLLSRRYGITAEQLDGLNSLGIHYDKERLNGKRLLEWEKLTGLDVRAIVAI

AMAESSLGTVGVAKEKGANMFGYGAFDFNPNNAKKYSDEVAIRHMVEDTIIANKNOTPERQDLKAKKNSL  
GQLDTLIDGGVYPTDTSGSGQRRADIMTKLDQWIDHGSTPEIPEHLKITSQTQFSEVPVGYKRSQPNV  
LTYKSBTYSFGQCTWYAYNRVKELGYQVDRYMGNGGDWQRKPGFVTTHKPKVGYVVSFAPGQAGADATYG  
HVAVVEQIKEDGSILISESNVMGLGTISYRTFTAEQASLLTYVVGDKLPRP

**SEQ ID NO: 52**

ATGAAGAAAAGGAAATTGTTAGCAGTAACACTATTAAGTACCATACTCTTAAACAGTGCAGTGCCATTAG  
TTGTTGCTGATACCTCCTTGCGTAATAGCACATCATCCACTGATCAGCCTACTACAGCAGATACTGATAC  
GGATGACGAGAGTGAAACACCAAAAAAGACAAAAAAGCAAGGAAACAGCGTCGCAGCACGACACCCAA  
AAAGACCATAAGCCATCACACACTCACCCAACCCCCCTTCAAATGATACTAAGCAGACCGATCAGGCAT  
CATCTGAAGCTACTGACAAACCAATAAAGACAAAAACGACACCAAGCAACCAGACAGCAGTGATCAATC  
CACCCCATCTCCCAAAGACCAGTCGTCTCAAAAAGAGTCACAAAACAAAGACGGCCGACCTACCCCATCA  
CCTGATCAGCAAAAAGATCAGACACCTGATAAAACACCAGAAAAATCAGCTGATAAAACCCCTGAAAAAG  
GACCAGAAAAAGCAACTGATAAAACACCAGAGCCAAATCGTGACGCTCCAAAACCCATCCAACCTCCTTT  
AGCAGCTGCTCCTGTCTTTATACCTTGGAGAGAAAGTGACAAAGACCTGAGCAAGCTAAAACCAAGCAGT  
CGCTCATCAGCGGCTTACGTGAGACACTGGACAGGTGACTCTGCCTACACTCACAACCTGTTGTACGCC  
GTTATGGGATTACTGCTGAACAGCTAGATGGTTTTTTGAACAGTCTAGGTATTCATATGATAAAGAACG  
CTTAAACGGAAGCGTTTATTAGAATGGGAAAACTAACAGGACTAGACGTTGAGCTATCGTAGCTATT  
GCAATGGCAGAAAGCTCACTAGGTACTCAGGGAGTTGCTAAAGAAAAAGGAGCCAATATGTTTGGTTATG  
GCGCCTTTGACTTCAACCCAAACAATGCCAAAAAATACAGCGATGAGGTTGCTATTCGTACATGGTAGA  
AGACACCATCATTGCCAACAAAAACCAACCTTTGAAAGACAAGACCTCAAAGCAAAAAAATGGTCACTA  
GGCCAGTTGGATACCTTGATTGATGGTGGGGTTTACTTTACAGATACAAGTGGCAGTGGGCAAAGACGAG  
CAGATATCATGACCAAACTAGACCAATGGATAGATGATCATGGAAGCACACCTGAGATTCCAGAACATCT  
CAAGATAACTTCCGGGACACAATTTAGCGAAGTGCCCGTAGGTTATAAAGAAGTCAGCCACAAAACGTT  
TTGACCTACAAGTCAGAGACCTACAGCTTTGGCCAATGCACTTGGTACGCCTATAATCGTGTCAAAGAGC  
TAGGTTATCAAGTCGACAGGTACATGGGTAACGGTGGCGACTGGCAGCGCAAGCCAGGTTTTGTGACCAC  
CCATAAACCTAAAGTGGGCTATGTCGTCTCATTTCACCAGGCCAAGCAGGAGCAGATGCAACCTATGGT  
CACGTTGCTGTTGTAGAGCAAATCAAAGAAGATGGTTCTATCTTAATTTACAGAGTCAAATGTTATGGGAC  
TAGGCACCATTTCTATCGGACGTTACAGCTGAGCAGGCTAGTTTGTGACCTATGTCGTAGGGGACAA  
ACTCCCAAGACCATAA

Preferred GAS 193 proteins for use with the invention comprise an amino acid sequence: (a) having  
50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 51; and/or (b) which is a fragment of at least  $n$   
consecutive amino acids of SEQ ID NO: 51, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 193 proteins include variants (e.g.  
allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 51. Preferred fragments  
of (b) comprise an epitope from SEQ ID NO: 51. Other preferred fragments lack one or more amino  
acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more  
amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID  
NO: 51. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide,  
of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

**(27) GAS 137**

GAS 137 corresponds to M1 GenBank accession numbers GI:13621842, GI:15674720 and  
GI:30173478, to M3 GenBank accession number GI:21909998, to M18 GenBank accession number  
GI: 19745749 and is also referred to as 'Spy0652' (M1), 'SpyM3\_0462', and 'SpyM18\_0713' (M18).  
Amino acid and polynucleotide sequences of GAS 137 of an M1 strain are set forth below:

**SEQ ID NO: 53**

MSDKHINLVIVTGMMSGAGKTVAIQSFEDLGYFTIDNMPPALVPKFLELIEQTNENRRVALVDMRSRLFF  
KEINSTLDSIESNPSIDFRILFLDATDGELVSRYKETRRSHPLAADGRVLDGIRLERELLSPLKSMSQHV  
VDTTKLTPRQLRKTISDQFSEGSNQASFRIEVMSEFGKYGLPLDADLVFDVRFLPNPYYQVELREKTGLD



EDVFNVMSEVSEVFPYKHLNLIIVPILPAYQKEGKSVLTVAIGCTGGQHRVAFHCLAESLATDWSVN  
BSHRDQNRKBTNRS

**SEQ ID NO: 54**

5 ATGTCAGACAAACACATTAATTTAGTTATTGTGACAGGAATGAGCGGCGCTGGAAAAACAGTTGCCATT  
AGTCTTTTGAGGATCTAGGCTACTTTACCATTGATAATATGCCCCAGCCTTGGTTCCAAAATTTTAGA  
ATTAATTGAACAAACCAATGAAAATCGTAGGGTGGCTTTGGTTGTCGATATGAGAAGTCGTTTGTTC  
AAGGAAATTAATCTACCTTAGATAGTATTGAAAGCAATCCTAGCATTGATTTTCGGATTCTTTTGG  
10 ATGCAACGGATGGAGAATTGGTGTACGCTATAAAGAAACCAGACGGAGCCACCCTTTGGCTGCGGACGG  
TCGTGTGCTTGATGGTATTTCGATTGAAAGAGAACTCCTATCTCCTTTGAAAAGCATGAGCCAACATGTG  
GTGGATACAACAAAATTGACCCCTAGACAATTGCGTAAAACCATTCAGACCAGTTTCTGAAGGGTCTA  
ATCAAGCCTCTTTTCGTATTGAAGTGATGAGCTTTGGGTTCAAATATGGTCTTCCTTTGGATGCGGATT  
GGTTTTTGATGTGCGTTTCTACCCAATCCTTATTATCAGGTAGAGCTTCGTGAAAAACAGGACTAGAT  
GAGGACGTTTTTAATTATGTGATGTCTACCCAGAATCAGAGGTGTTTTACAAGCATTGTAAACCTTA  
15 TTGTCCTATCTTACCGGCTTACCAAAAAGAAGGGAAGTCTGTCTTGACGGTGGCTATTGGCTGCACAGG  
AGGCCAACACCGCAGCGTTGCCTTTGCCATTGCTTGGCAGAAAGTCTGGCAACAGATTGGTCGGTTAAT  
GAAAGCCATCGTGATCAAAATCGTCGTAAGGAAACGGTGAATCGTTCATGA

Preferred GAS 137 proteins for use with the invention comprise an amino acid sequence: (a) having  
20 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 53; and/or (b) which is a fragment of at least *n*  
consecutive amino acids of SEQ ID NO: 53, wherein *n* is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 137 proteins include variants (e.g.  
allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 53. Preferred fragments  
25 of (b) comprise an epitope from SEQ ID NO: 53. Other preferred fragments lack one or more amino  
acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more  
amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID  
NO: 53. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide,  
of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

30 **(28) GAS 084**

GAS 084 corresponds to M1 GenBank accession numbers GI:13622398 and GI:15675229, to M3  
GenBank accession number GI: 21910442, to M18 GenBank accession number GI: 19746199 and is  
also referred to as 'Spy1274' (M1), 'SpyM3\_0906' and 'SpyM18\_1223' (M18). GAS 084 has also  
been identified as a putative amino acid ABC transporter/periplasmic amino acid binding protein.

35 Amino acid and polynucleotide sequences of GAS 084 of an M1 strain are set forth below:

**SEQ ID NO: 55**

40 MIKKRTVAIALAIASSFFLVACQATKSLKSGDAWGVYQKQKSITVGFNDTFVPMGYKDESGRCKGFDIDL  
AKEVPHQYGLKVNQAINWDMKEAELNNGKIDVIWNGYSITKERQDKVAFTDSYMRNEQIIIVKKRSDIK  
TISDMKHKVLGAQSASSGYDSLRLTPKLLKDFIKNKDANQYETFTQAFIDLKSDRIDGILIDKVYANYYL  
AKEGQLENYRMIPTTFENEAFSVGLRKEDKTLQAKINRAFRVLYQNGKFQAISEKWFGDDVATANIKS

**SEQ ID NO: 55**

45 ATGATTATAAAAAAAGAACCGTAGCAATTTAGCCATAGCTAGTAGCTTTTCTTGGTAGCTTGTCAG  
CTACTAAAGTCTTAAATCAGGAGATGCTTGGGGAGTTTACCAAAAGCAAAAAGTATTACAGTTGGTTT  
TGACAATACGTTTGTTCCTATGGGCTATAAGGATGAAAGCGGCAGATGCAAAGGTTTGTATATTGATTG  
GCTAAAGAAGTTTTTACCAATATGGAATCAAGGTTAACTTTCAAGCTATTAATTGGGACATGAAAGAAG  
CAGAACTAAACAATGGTAAATGATGTAATCTGGAATGGTTATTCAATAACTAAGGAGCGTCAGGATAA  
GGTTGCCTTTACTGATTCTTACATGAGAAATGAACAAATTATTGTTGTCAAAAAAAGATCTGATATAAA  
50 ACAATATCAGATATGAAACATAAAGTGTTAGGAGCACAATCAGCTTCATCAGGTTATGACTCCTTGTTAA  
GAACTCCTAAACTGCTGAAAGATTTTATTAATAAATAAAGACGCTAATCAATATGAAACCTTTACACAAGC

TTTTATTGATTAAATCAGATCGTATCGATGGAATATTGATTGACAAAGTATATGCCAATTACTATTTA  
GCAAAAGAAGGGCAATTAGAGAATTATCGGATGATCCCAACGACCTTTGAAAATGAAGCATTTCGGTTG  
GACTTAGAAAAGAAGACAAAACGTTGCAAGCAAAAATTAATCGTGCTTTCAGGGTGCTTTATCAAAATGG  
CAAATTTCAAGCTATTTCTGAGAAATGTTTGGAGATGATGTTGCCACTGCCAATATTAAATCTTAA

5

Preferred GAS 084 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 55; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 55, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 084 proteins include variants (e.g. allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 55. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 55. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 55. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 55 is removed. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

10

15

#### (29) GAS 384

GAS 384 corresponds to M1 GenBank accession numbers GI:13622908 and GI:15675693, to M3 GenBank accession number GI: 21911154, to M18 GenBank accession number GI: 19746801 and is also referred to as 'Spy1874' (M1), 'SpyM3\_1618' (M3), and 'SpyM18\_1939' (M18). GAS 384 has also been identified as a putative glycoprotein endopeptidase. Amino acid and polynucleotide sequences of GAS 384 of an M1 strain are set forth below:

20

#### SEQ ID NO: 57

25

MKTLAFDTSNKTLNLAILDDETLADMTLNIQKKHSVSLMPAIDFLMTCTDLKPQDLERIVVAKGPGSYT  
GLRVAVATAKTLAYSLNIALVGISSLYALAASTCKQYFNTLVVPLIDARRQNAYVGYRQKSVMPQAH  
SLEVIIEQLVEEGQLIFVGETAPFAEKIQKKLPQAILLPTLPSAYECGLLGQSLAPENVDAFVPQYLKRV  
EAEENWLKDNEIKDDSHYVKRI

30

#### SEQ ID NO: 58

35

ATGAAGACACTTGCAATTTGATACCTCAAATAAAACCTTGTCCTTGCTATACTTGATGATGAGACACTTC  
TAGCAGATATGACCCTTAACATTCAGAAAAACATAGTGTTAGCCTTATGCCTGCTATTGATTTTTTGT  
GACTTGTAAGTATCTTAAACCTCAAGATTTAGAAAGAATAGTGGTTGCAAAGGCCCTGGATCTTACACA  
GGTTTACGAGTGGCAGTTGCTACTGCAAAAACGTTAGCGTACAGTTTAAATATTGCATTGGTCGGGATTT  
CGAGTCTATATGCTTTGGCTGCGTCTACTTGTAACAGTATCCAAATACTTTGGTGGTGCCATTGATTGA  
TGCTAGAAGGCAAAATGCGTATGTAGGTTATTATCGGCAAGGAAAATCAGTGATGCCACAAGCCCATGCT  
TCACTAGAAGTTATTATAGAACAATTAGTAGAAGAAGGACAGCTGATTTTTGTTGGGGAGACTGCTCCTT  
TTGCTGAGAAAATTCAAAGAACTACCTCAGGCGATACTACTTCCAACCCTTCTTCTGCTTACGAATG  
TGGTCTTTTGGGGCAAAGTTTGGCACCAGAAAATGTAGACGCCTTTGTCCCTCAATATCTCAAGAGAGTG  
GAAGCTGAAGAAAACCTGGCTCAAAGATAATGAGATAAAAGATGATAGTCACTACGTTAAGCGAATCTAA

40

Preferred GAS 384 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 57; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 57, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 384 proteins include variants (e.g. allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 57. Preferred fragments

45



of (b) comprise an epitope from SEQ ID NO: 57. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 57. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

### (30) GAS 202

GAS 202 corresponds to M1 GenBank accession numbers GI:13622431 and GI:15675258, to M3 GenBank accession number GI: 21910527, to M18 GenBank accession number GI: 19746290 and is also referred to as 'Spy1309' (M1), 'SpyM3\_0991' (M3), 'SpyM18\_1321' (M18) and 'dltD'. GAS 202 has also been identified as a putative extramembranal protein. Amino acid and polynucleotide sequences of GAS 202 of an M1 strain are set forth below:

#### SEQ ID NO: 59

MLKRLWLIIGPLLI AFVLVVITIFSFPTQLDHSIAQEKANAVAITDSSFKNGLIKRQALSDETCRFVPPF  
GSSEWSRMDSMHPSVLAERYKRSYRPFLLIGKRGSA SLSHYYGIQQITNEMQKKKAI FVVSPQWFTAQGIN  
PSAVQMYLSNTQVIEFLLKARTDKESQFAAKRLLLELNPVSKSNLLKKVSKGKSLSRDLRAILKQCQHQVA  
LREESLFSFLGKSTNYEKRI LPRVKGLPKVFSYKQLNALATKRGQLATTNNRFGIKNTFYRKRIAPKYNL  
YKNFQVNYSYLASPEYNDFQLLLSEFAKRKTDVLFVITPVNKAWADYTGLNQDKYQAAVRKIKFQLKSQG  
FHRIADFSKGGESYFMQDTIHLGWNGWLAFDCKVQPFLETKQVPVNYKMPYFYSKIWANRKDLQ

#### SEQ ID NO: 60

ATGCTTAAGAGACTCTGGTTAATTCTAGGTCTCTTCTTATTGCCTTTGTTTTAGTAGTGATTACTATTT  
TTAGTTTTCTACACAACCTTGATCATTCCATAGCTCAGGAAAAAGCAAATGCCGTGCGATCACAGATAG  
TTCTTTTAAAAATGGTTTGATTAAAAGACAAGCTTTATCAGATGAGACTTGTCGTTTTGTGCCTTTTTTT  
GGTTCTAGCGAATGGAGTCGAATGGATAGTATGCACCCTTCGGTGCTTGAGAGCGCTACAAGCGGAGCT  
ATAGACCATTTTTAATTGGTAAGAGAGGATCAGCATCTTTGTGCGATTATTATGGTATACAACAAATTAC  
CAATGAAATGCAAAAGAAAAAGCCATCTTTGTAGTATCTCCTCAATGGTTTACTGCTCAAGGGATTAAT  
CCTAGTGCGGTTTCAGATGTACTTGTCTAACACTCAAGTGATTGAATTTTTACTAAAAGCTAGAACTGATA  
AAGAATCACAGTTTGCAGCAAAGCGTTTGCTTGAGCTTAACCCCTGGTGTGTCTAAATCAAACCTTATTGAA  
AAAAGTAAGTAAGGGTAAGTCTCTTAGTCGGTTAGACAGAGCTATTTTGAAATGTCAACATCAAGTAGCA  
TTGAGAGAAGAGTCCCTTTTGTAGTTTTTATAGGCAAATCTACTAATACTATGAAAAAGAATTTTGCTCGCG  
TTAAGGGATTACCTAAAGTATTTTCGTATAACAATTGAATGCATTAGCAACTAAGAGAGGCCAATTAGC  
AACAACCAACAACCGTTTTGGGATTAAAAATACATTTTATCGTAAACGAATAGCACCTAAATACAATCTT  
TATAAGAAATTTCCAAGTTAATTATAGTTACCTGGCGTCACCAGAATACAATGATTTTCAGCTTTTATTAT  
CAGAATTTGCTAAACGAAAAACAGATGTACTCTTTGTTATAACTCTGTTAATAAAGCTTGCGCGGATTA  
TACCGGCTTAAATCAAGATAAGTATCAAGCGGCAGTTCGTAAATAAAATTCCAGTTAAAGTCACAAGGA  
TTTCATCGCATTTGCTGACTTCTCAAAAGATGGTGGTGAGTCTTACTTTATGCAAGATACCATCCATCTCG  
GTTGGAATGGCTGGTTAGCTTTTGATAAGAAAGTGCAACCATTCTAGAAACGAAGCAGCCAGTGCCCAA  
CTATAAAATGAACCCTTATTTTTATAGTAAATTTGGGCAAATAGGAAAGACTTGCAATAG

Preferred GAS 202 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 59; and/or (b) which is a fragment of at least  $n$  consecutive amino acids of SEQ ID NO: 59, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 202 proteins include variants (e.g. allelic variants, homologs, orthologs, paralog, mutants, etc.) of SEQ ID NO: 59. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 59. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID

NO: 59. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

**(31) GAS 057**

GAS 057 corresponds to M1 GenBank accession numbers GI:13621655 and GI:15674549, to M3  
5 GenBank accession number GI: 21909834, to M18 GenBank accession number GI: 19745560 and is  
also referred to as 'Spy0416' (M1), 'SpyM3\_0298' (M3), 'SpyM18\_0464' (M18) and 'prtS'. GAS  
057 has also been identified as a putative cell envelope proteinase. Amino acid and polynucleotide  
sequences of GAS 057 of an M1 strain are set forth below:

**SEQ ID NO: 61**

10 MEKKQRFSLRKYKSGTFSVLIGSVFLVMTTVADELSTMSEPTITNHAQQQAQHLTNTELSSAESKSQD  
TSQITLKTNRKEQSQDLVSEPTTTELADTDAASMAN TGSDATQKSASLPPVNTDVHDVVKTKGAWDKGY  
KGQGVVAVIDTGIDPAHQSMRISDVSTAKVKSKEMLARQKAAGINYGSWINDKVVFHNYVENS DN IK  
ENQFEDFDEDWENFBFDAEAPKAIKKHKIYRPQSTQAPKETVIKTEETDGSHDIDWTQTD DDTKYESHG  
15 MHVTGIVAGNSKEAAATGERFLGIAPEAQVMFMRV FANDIMGSAESLFIKAI EDAVALGADV INLSLGT A  
NGAQLSGSKPLMBAIEKAKKAGVSVVVAAGNERVYGS DHDDPLATNP DYGLV GSPSTGRTP TSVAAINS K  
WVIQRLMTVKELENRADLNHGKAIYSESVDPKDIKDSLGYDKSHQFAYVKESTDAGYNAQDVKGKIALIE  
RDPNKTYDEMI ALAKKHGALGVLI FNNKPGQSNRSMRLTANGMGI PSAFI SHEFGKAMSQ L N G N T G S L E  
FDSVVS KAPSQKGNEMNHFSNWGLTSDGYLKPDI TAPGGDIYSTYNDNHYGSQTGTSMAS PQIAGASLLV  
20 KQYLEKTQPNLPKEKIADIVKNLLMSNAQIHVNPETKTTTSPRQQGAGLLNIDGAVTSGLYVTGKDN YGS  
ISLGNITDTMTFDVTVHNLSNKDKTLRYDT ELLTDHVDPQKGRFTLTSHSLKTYQGGEVTV PANGKVTVR  
VTMDVSQFTKELTKQMPNGYYLEGFVRFRDSQDDQLNRVNI PFVGFKGQFENLAVA EESIYRLKSQKGTG  
FYFDES GPKDDIYVGKHFTGLVTLGSETNVSTKTI SDNGLHTLGT FKNADGKFILEKNAQGNPVLAI SPN  
GDNNQDFAAFKGVFLRKYQGLKASVYHASDKEHKNPLWVSPESFKGDKNFNSDIRFAKSTLLGTAFSGK  
SLTGAELPDGHYHYVVSYPDVVGAKRQEMTFDMI LDRQKPVLSQATFDPETNRFKPEPLKDRGLAGVRK  
25 DSVFYLERKDNKPYTVTINDSYKYVSVEDNKT FVERQADGSFILPLDKAKLGDFYIMVEDFAGNVAIAKL  
GDHLPQTLGKTPIK LKLTGNYQTKETLKD NLEMTQSDTGLVTNQAQLAVVHRNQPSQLTKMNQDFFIS  
PNEDGNKDFVAFKGLKNNVYNDLTVNVYAKDDHQKQTP IWSSQAGASVSAIESTAWYGITARGSKVMPGD  
YQYVVTYRDEHGKEHQKYTISVNDKPKMITQGRFD TINGVDHFTPDKTKALDSSGIVREEVFYLA KKNG  
RKFDVTEGKDGITVSDNKVYIPKNPDGSYTI SKRDGVTLS DYYYLVEDRAGNVSFATLRDLKAVGKDKAV  
30 VNFGLDLPVPEDKQIVNFTYLVRDADGKPIENLEYNNSGNSLILPYGKYTV ELLTYDTNAAKLES DKIV  
SFTLSADNNFQQVTFKITMLATSQITAHFDHLLPEGS RVSLKTAQDQLI PLEQSLYVPKAYGKTVQEGTY  
EVVVSLPKGYRIE GNTKVNTLPNEVHEL SLRLVKVG DASDSTGDHKVMSKNNSQAL TASATPTKSTTSAT  
AKALPSTGEKMG LKLRIVGLVLLGLTCVFSRKKSTKD

**SEQ ID NO: 62**

35 GTGGAGAAAAGCAACGTTTTTCCCTTAGAAAATACAAATCAGGAACGTTTTTCGGTCTTAATAGGAAGCG  
TTTTCTTGGTGATGACAACAACAGTAGCAGCAGATGAGCTAAGCACAATGAGCGAACCAACAATCACGAA  
TCACGCTCAACAACAAGCGCAACATCTCACC AATACAGAGTTGAGCTCAGCTGAATCAAAATCTCAAGAC  
ACATCACA AATCACTCTCAAGACAAATCGTGAAAAAGAGCAATCACAAGATCTAGTCTCTGAGCCAACCA  
40 CAACTGAGCTAGCTGACACAGATGCAGCATCAATGGCTAATACAGGTTCTGATGCGACTCAAAAAGCGC  
TTCTTTACCGCCAGTCAATACAGATGTTACGATTGGGTAAAAACCAAGGAGCTTGGGACAAGGGATAC  
AAAGGACAAGGCAAGGTTGTGCGAGTTATTGACACAGGGATCGATCCGGCCCATCAAAGCATGCGCATCA  
GTGATGTATCAACTGCTAAAGTAAAATCAAAAGAAGACATGCTAGCACGCCAAAAGCCGCCGCTATTAA  
TTATGGGAGTTGGATAAATGATAAAGTTGTTTTTGACATAATTATGTGGAAAATAGCGATAATATCAAA  
45 GAAAATCAATTCGAGGATTTTGATGAGGACTGGGAAAAC TTTGAGTTTGATGCAGAGGCAGAGCCAAAAG  
CCATCAAAAACACAAGATCTATCGTCCCCAATCAACCCAGGCACCGAAAGAACTGTTATCAAAACAGA  
AGAAACAGATGGTTCATGATATTGACTGGACACAAACAGACGATGACACCAAATACGAGTCACACGGT  
ATGCATGTGACAGGTATTGTAGCCGGTAATAGCAAAGAAGCCGCTGCTACTGGAGAACGCTTTTTAGGAA  
TTGCACCAGAGGCCCAAGTCATGTTTCATGCGTGT TTTTGCCAACGACATCATGGGATCAGCTGAATCACT  
50 CTTTATCAAAGCTATCGAAGATGCCGTGGCTTTAGGAGCAGATGTGATCAACCTGAGTCTTGGAACCGCT  
AATGGGGCAGCCTTAGTGGCAGCAAGCCTCTAATGGAAGCAATTGAAAAAGCTAAAAAGCCGGTGTAT  
CAGTTGTTGTAGCAGCAGGAAATGAGCGGTCTATGGATCTGACCATGATGATCCATTGGCGACAAATCC  
AGACTATGGTTTGGTCGGTTCTCCCTCAACAGGTGGAACACCAACATCAGTGGCAGCTATAAACAGTAAG  
TGGGTGATTCAACGTCTAATGACGGTCAAAGAATTAGAAAACCGTGCCGATTTAAACCATGGTAAAGCCA  
55 TCTATT CAGAGTCTGTGCACTTTAAAGACATAAAAGATAGCCTAGGTTATGATAAATCGCATCAATTTGC



TTATGTCAAAGAGTCAACTGATGCGGGTTATAACGCACAAGACGTTAAAGGTAATGCTTTAATTGAA  
 CGTGATCCCAATAAAACCTATGACGAAATGATTGCTTTGGCTAAGAAACATGGAGCTCTGGGAGTACTTA  
 TTTTAAATAACAAGCCTGGTCAATCAAACCGCTCAATGCGTCTAACAGCTAATGGGATGGGGATACCATC  
 5 TGCTTTTCATATCGCACGAATTTGGTAAGGCCATGTCCCAATTAAATGGCAATGGTACAGGAAGTTTAGAG  
 TTTGACAGTGTGGTCTCAAAGCACCAGTCAAAAAGGCAATGAAATGAATCATTTTTCAAATTGGGGCC  
 TAACTTCTGATGGCTATTTAAAACCTGACATTACTGCACCAGGTGGCGATATCTATTCTACCTATAACGA  
 TAACCACTATGGTAGCCAAACAGGAACAAGTATGGCCTCTCCTCAGATTGCTGGCGCCAGCCTTTTGGTC  
 AAACAATACCTAGAAAAGACTCAGCCAAACTTGCCAAAAGAAAAAATTGCTGATATCGTTAAGAACCTAT  
 10 TGATGAGCAATGCTCAAATTCATGTTAATCCAGAGACAAAAACGACCACCTCACCGCGTCAGCAAGGGGC  
 AGGATTACTTAATATTGACGGAGCTGTCACTAGCGGCCTTTATGTGACAGGAAAAGACAACCTATGGCAGT  
 ATATCATTAGGCAACATCACAGATACGATGACGTTTGATGTGACTGTTTACAACCTAAGCAATAAAGACA  
 AAACATTACGTTATGACACAGAATTGCTAACAGATCATGTAGACCCACAAAAGGGCCGCTTCACTTTGAC  
 TTCTCACTCCTTAAAAACGTACCAAGGAGGAGAAGTTACAGTCCAGCCAATGGAAAAGTGACTGTAAGG  
 GTTACCATGGATGTCTCACAGTTCACAAAAGAGCTAACAAAACAGATGCCAAATGGTTACTATCTAGAAG  
 15 GTTTTGTCCGCTTTAGAGATAGTCAAGATGACCAACTAAATAGAGTAAACATTCTTTTGTGGTTTAA  
 AGGGCAATTTGAAAACCTTAGCAGTTGCAGAAGAGTCCATTACAGATTAAATCTCAAGGCAAACTGGT  
 TTTTACTTTGATGAATCAGGTCCAAAAGACGATATCTATGTCCGTAAACACTTTACAGGACTTGTCACCTC  
 TTGGTTCAGAGACCAATGTGTCAACCAAAACGATTTCTGACAATGGTCTACACACACTTGGCACCTTTAA  
 AAATGCAGATGGCAAATTTATCTTAGAAAAAATGCCCAAGGAAACCTGTCTTAGCCATTTCTCCAAAT  
 20 GGTGACAACAACCAAGATTTTGCAGCCTTCAAAGGTGTTTCTTGAGAAAATATCAAGGCTTAAAGCAA  
 GTGTCTACCATGCTAGTGACAAGGAACACAAAAATCCACTGTGGGTGAGCCAGAAAGCTTTAAAGGAGA  
 TAAAAACTTTAATAGTGACATTAGATTTGCAAAATCAACGACCCTGTTAGGCACAGCATTTTCTGGAAAA  
 TCGTTAACAGGAGCTGAATTACCAGATGGGCATTATCATTATGTGGTGTCTTATTACCCAGATGTGGTCG  
 GTGCCAAACGTCAAGAAATGACATTTGACATGATTTTAGACCGACAAAAACCGGTACTATCACAAGCAAC  
 25 ATTTGATCCTGAAACAAACCGATTCAAACCAGAACCCCTAAAGACCGTGGATTAGCTGGTGTTCGCAA  
 GACAGTGTCTTTTATCTAGAAAGAAAAGACAACAAGCCTTATACAGTTACGATAAACGATAGCTACAAAT  
 ATGTCTCAGTAGAAGACAATAAAACATTTGTGGAGCGACAAGCTGATGGCAGCTTTATCTTGCCGCTTGA  
 TAAAGCAAATTAGGGGATTTCTATTACATGGTTCGAGGATTTTGCAGGGAACGTGGCCATCGCTAAGTTA  
 GGAGATCACTTACCACAAACATTAGGTAAAACACCAATTAACTTAAGCTTACAGACGGTAATTATCAGA  
 30 CCAAAGAAACGCTTAAAGATAATCTTGAAATGACACAGTCTGACACAGGTCTAGTCACAAATCAAGCCCA  
 GCTAGCAGTGGTGACCGCAATCAGCCGCAAAGCCAGCTAACAAAGATGAATCAGGATTTCTTTATCTCA  
 CCAAACGAAGATGGGAATAAAGACTTTGTGGCCTTTAAAGGCTTGAAAAATAACGTGTATAATGACTTAA  
 CGGTTAACGTATACGCTAAAGATGACCACCAAAAACAAACCCCTATCTGGTCTAGTCAAGCAGGCGCTAG  
 TGTATCCGCTATTGAAAGTACAGCCTGGTATGGCATAACAGCCCGAGGAAGCAAGGTGATGCCAGGTGAT  
 35 TATCAGTATGTTGTGACCTATCGTGACGAACATGGTAAAGAACATCAAAGCAGTACACCATATCTGTGA  
 ATGACAAAAACCAATGATCACTCAGGGACGTTTGTATACCATTAAATGGCGTTGACCACTTTACTCCTGA  
 CAAGACAAAAGCCCTTGACTCATCAGGCATTGTCCGCGAAGAAGTCTTTTACTTGGCCAAGAAAAATGGC  
 CGTAAATTTGATGTGACAGAAGGTAAAGATGGTATCACAGTTAGTGACAATAAGGTGTATATCCCTAAAA  
 ATCCAGATGGTTCTTACACCATTTCAAAGAGAGATGGTGTACACTGTGAGATTATTACTACCTTGTCGA  
 40 AGATAGAGCTGGTAATGTGTCTTTTGCTACCTTGCGTGACCTAAAGCGGTGCGAAAAGACAAAGCAGTA  
 GTCAATTTTGGATTAGACTTACCGGTCCCTGAAGACAAACAAATAGTGAACCTTACCTACCTTGTCGGG  
 ATGCAGATGGTAAACCGATTGAAAACCTAGAGTATTATAAATACTCAGGTAACAGTCTTATCTTGCCATA  
 CGGCAAATACAGGTGCAATTGTTGACCTATGACACCAATGCAGCCAAACTAGAGTCAGATAAAATCGTT  
 TCCTTTACCTTGTCAGCTGATAACAACCTCCAACAAGTTACCTTTAAGATAACGATGTTAGCAACTTCTC  
 45 AAATAACTGCCCACTTTGATCATCTTTTGCCAGAAGGCAGTTCGCGTTAGCCTTAAAAACAGCTCAAGATCA  
 GCTAATCCCGCTTGAACAGTCTTGTATGTGCCTAAAGCTTATGGCAAACCGTTCAAGAAGGCACTTAC  
 GAAGTTGTTGTGACGCTGCCTAAAGGCTACCGTATCGAAGGCAACACAAAGGTGAATACCCTACCAATG  
 AAGTGACGAACTATCATTACGCTTGTCAAAGTAGGAGATGCCTCAGATTCAACTGGTGTATCATAAGGT  
 TATGTCAAAAAATAATTCACAGGCTTTGACAGCCTCTGCCACACCAACCAAGTCAACGACCTCAGCAACA  
 50 GCAAAAGCCCTACCATCAACGGGTGAAAAAATGGGTCTCAAGTTGCGCATAGTAGGTCTTGTTGTTACTCG  
 GACTTACTTGCGTCTTTAGCCGAAAAAATCAACCAAAGATTGA

Preferred GAS 057 proteins for use with the invention comprise an amino acid sequence: (a) having  
 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
 55 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 61; and/or (b) which is a fragment of at least  $n$   
 consecutive amino acids of SEQ ID NO: 61, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or more). These GAS 057 proteins include variants (e.g.  
 allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 61. Preferred fragments

of (b) comprise an epitope from SEQ ID NO: 61. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 61. For example, in one embodiment, the underlined amino acid sequence at the N-terminus of SEQ ID NO: 61 is removed. In another example, the underlined amino acid sequence at the C-terminus of SEQ ID NO: 61 is removed. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

The immunogenicity of other known GAS antigens may be improved by combination with two or more GAS the first antigen group. Such other known GAS antigens include a second antigen group consisting of (1) one or more variants of the M surface protein or fragments thereof, (2) fibronectin-binding protein, (3) streptococcal heme-associated protein, or (4) SagA. These antigens are referred to herein as the "second antigen group".

The invention thus includes an immunogenic composition comprising a combination of GAS antigens, said combination consisting of two to thirty-one GAS antigens of the first antigen group and one, two, three, or four GAS antigens of the second antigen group. Preferably, the combination consists of three, four, five, six, seven, eight, nine, or ten GAS antigens from the first antigen group. Still more preferably, the combination consists of three, four or five GAS antigens from the first antigen group. Preferably, the combination of GAS antigens includes either or both of GAS 40 and GAS 117. Preferably, the combination of GAS antigens includes one or more variants of the M surface protein.

Each of the GAS antigens of the second antigen group are described in more detail below.

*(1) M surface protein*

Over 100 different type variants of the M protein have been identified. Epitopes having increased bactericidal activity and having decreased likelihood of cross-reacting with human tissues have been identified in the amino terminal region and combined into fusion proteins containing approximately six, seven, or eight M protein fragments linked in tandem. See Ref. 4, 5, 6, WO 02/094851 and WO 94/06465. (Each of the M protein variants, fragments and fusion proteins described in these references are specifically incorporated herein by reference.)

Accordingly, the compositions of the invention may further comprise a GAS M surface protein or a fragment or derivative thereof. One or more GAS M surface protein fragments may be combined together in a fusion protein. Alternatively, one or more GAS M surface protein fragments are combined with a GAS antigen or fragment thereof of the first antigen group. One example of a GAS M protein is set forth below.

**SEQ ID NO: 63**

MAKNTNRHYSRLKLTGTASVAVALTVLGAGFANQTEVKANGDGNPREVIEDLAANNPAIQNIRLRYEN  
KDLKARLENAMEVAGRDFKRABELEKAKQALEDRKDLKLELQDYDLAKESTSWDRQRLEKELEBK



KEALELAIDQASRDYHRATALEKEBLEBKKALELAIDQASQDYNRANVLEKELETITREQBINRNLLGNA  
 KLBDQLSSEKEQLTIEKAKLEBBKQISDASRQSLRRDLASREAKKQVEKDLANLTAELDKVKEDKQIS  
 DASRQGLRRDLASREAKKQVEKDLANLTAELDKVKEEKQISDASRQGLRRDLASREAKKQVEKALEBA  
 NSKLALEKLNKELESKLTKEKBAELQAKLEBAKALKEQLAKQABBLAKLRAGKASDSQTPDTKPGN  
 5 KAVPGKGQAPQAGTKPNQNKAPMKETKRQLPSTGETANPPPTAAALTMATAGVAAVVKRKEEN

Preferred GAS M proteins for use with the invention comprise an amino acid sequence: (a) having  
 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 63; and/or (b) which is a fragment of at least  $n$   
 10 consecutive amino acids of SEQ ID NO: 63, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
 30, 35, 40, 50, 60, 70, 80, 90, 100, 150 or more). These GAS M proteins include variants (e.g. allelic  
 variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 63. Preferred fragments of (b)  
 comprise an epitope from SEQ ID NO: 63. Preferably, the fragment is one of those described in the  
 references above. Preferably, the fragment is constructed in a fusion protein with one or more  
 15 additional M protein fragments. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3,  
 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2,  
 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 63. Other fragments  
 omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain,  
 of a transmembrane domain, or of an extracellular domain).

## 20 (2) Fibronectin-binding protein

GAS fibronectin-binding protein ('Sfbl') is a multifunctional bacterial protein thought to mediate  
 attachment of the bacteria to host cells, facilitate bacterial internalization into cells and to bind to the  
 Fc fragment of human IgG, thus interfering with Fc-receptor mediated phagocytosis and antibody-  
 dependent cell cytotoxicity. Immunization of mice with Sfbl and an 'H12 fragment' (encoded by  
 25 positions 1240 – 1854 of the Sfbl gene) are discussed in Refs. 7,8 and 9. One example of an amino  
 acid sequence for GAS Sfbl is show below.

### SEQ ID NO: 64

MSFDGFFLHHLTNELKENLLYGRIQKVNQPFERELVLTIRNHRKNYKLLLSAHPVFGRVQITQADFQNPQ  
 VPNTFTMIMRKYLQGAVIEQLEQIDNDRIIEIKVSNKNEIGDAIQATLIEIMGKHSNIIIVDRAENKII  
 30 ESIKHVGFSSQNSYRTILPGSTYIEPPKTAAVNPFTITDVPLFEILQTQELTVKSLQQHFQGLGRDTAKEL  
 AELLTTDKLKRFPREFFARPTQANLTTASFPVLFSDSHATFETLSMDLDFYQDKAERDRINQQASDLIH  
 RVQTELDKNRNKLSKQEAELLATENAELFRQKGEELLTYLSLVPNNQDSVILDNYTGEKIEIALDKALT  
 PNQNAQRYFKKYQKLKEAVKHLGLIADTKQSITYFESVDYNLSQASIDDIEDIREELYQAGPLKSRQRD  
 KRHKRKKPEQYLASDGTIILMVGRNNLQNEELTFKMAKKGELWFAKDI PGSHV I I KDNLDPSDEVKTD A  
 35 AELAAYYSKARLSNLVQVDMIEAKKLHKPSGAKPGFVTYTGQKTLRVTPDQAKILSMKLS

Preferred Sfbl proteins for use with the invention comprise an amino acid sequence: (a) having 50%  
 or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,  
 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 64; and/or (b) which is a fragment of at least  $n$   
 40 consecutive amino acids of SEQ ID NO: 64, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25,  
 30, 35, 40, 50, 60, 70, 80, 90, 100, or more). These Sfbl proteins include variants (e.g. allelic variants,  
 homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 64. Preferred fragments of (b) comprise  
 an epitope from SEQ ID NO: 64. Preferably, the fragment is one of those described in the references  
 above. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15,

20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 64. Other fragments omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

### 5 (3) *Streptococcal heme-associated protein*

The GAS streptococcal heme-associated protein ('Shp') has been identified as a GAS cell surface protein. It is thought to be cotranscribed with genes encoding homologues of an ABC transporter involved in iron uptake in gram-negative bacteria. The Shp protein is further described in 10. One example of a Shp protein is shown below:

#### 10 SEQ ID NO: 65

MTKVVIKQLLQVIVVFMISLSTMTNLVYADKGQIYGCI IQRNYRHPISGQIEDSGGEHSFDIGQGMVEGT  
VYSDAMLEVSDAGKIVLTFRMSLADYSGNYQFWIQPGGTGSFQAVDYNITQKGTDTNGTTLDAISLPTV  
NSIIRGSMFVEPMGREVVFYLSASELIQKYSGNMLAQLVTETDNSQNQEVKDSQKPVDTKLGESQDESHT  
15 GAMITQNKPKANSNNKSLSDKKILPSKMGLTTSLELKKEDKFRSKDLSIMIYYFPTFFLMLGGFAVWV  
WKKRKKNDKTM

Preferred Shp proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 65; and/or (b) which is a fragment of at least  $n$   
20 consecutive amino acids of SEQ ID NO: 65, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 or more). These Shp proteins include variants (e.g. allelic variants, homologs, orthologs, paralog, mutants, etc.) of SEQ ID NO: 65. Preferred fragments of (b) comprise an epitope from SEQ ID NO: 65. Other preferred fragments lack one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the C-terminus and/or one or more amino acids (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more) from the N-terminus of SEQ ID NO: 65. Other fragments  
25 omit one or more domains of the protein (e.g. omission of a signal peptide, of a cytoplasmic domain, of a transmembrane domain, or of an extracellular domain).

### (4) *SagA*

Streptolysin S (SLS), also known as 'SagA', is thought to be produced by almost all GAS colonies.  
30 This cytolytic toxin is responsible for the beta-hemolysis surrounding colonies of GAS grown on blood agar and is thought to be associated with virulence. While the full SagA peptide has not been shown to be immunogenic, a fragment of amino acids 10 – 30 (SagA 10 – 30) has been used to produce neutralizing antibodies. See Ref. 11. The amino acid sequence of SagA 10 – 30 is shown below:

#### 35 SEQ ID NO: 66 FSLATGSGNSQGGSGSYTPGKC

Preferred SagA 10-30 proteins for use with the invention comprise an amino acid sequence: (a) having 50% or more identity (e.g. 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5% or more) to SEQ ID NO: 66; and/or (b) which is a fragment of at



least  $n$  consecutive amino acids of SEQ ID NO: 66, wherein  $n$  is 7 or more (e.g. 8, 10, 12, 14, 16, 18, or 20). These SagA 10 - 30 proteins include variants (e.g. allelic variants, homologs, orthologs, paralogs, mutants, etc.) of SEQ ID NO: 66.

5 There is an upper limit to the number of GAS antigens which will be in the compositions of the invention. Preferably, the number of GAS antigens in a composition of the invention is less than 20, less than 19, less than 18, less than 17, less than 16, less than 15, less than 14, less than 13, less than 12, less than 11, less than 10, less than 9, less than 8, less than 7, less than 6, less than 5, less than 4, or less than 3. Still more preferably, the number of GAS antigens in a composition of the invention is less than 6, less than 5, or less than 4. Still more preferably, the number of GAS antigens in a  
10 composition of the invention is 3.

The GAS antigens used in the invention are preferably isolated, i.e., separate and discrete, from the whole organism with which the molecule is found in nature or, when the polynucleotide or polypeptide is not found in nature, is sufficiently free of other biological macromolecules so that the polynucleotide or polypeptide can be used for its intended purpose.

#### 15 *Fusion proteins*

The GAS antigens used in the invention may be present in the composition as individual separate polypeptides, but it is preferred that at least two (i.e. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20) of the antigens are expressed as a single polypeptide chain (a 'hybrid' polypeptide). Hybrid polypeptides offer two principal advantages: first, a polypeptide that may be unstable or  
20 poorly expressed on its own can be assisted by adding a suitable hybrid partner that overcomes the problem; second, commercial manufacture is simplified as only one expression and purification need be employed in order to produce two polypeptides which are both antigenically useful.

The hybrid polypeptide may comprise two or more polypeptide sequences from the first antigen group. Accordingly, the invention includes a composition comprising a first amino acid sequence and  
25 a second amino acid sequence, wherein said first and second amino acid sequences are selected from a GAS antigen or a fragment thereof of the first antigen group. Preferably, the first and second amino acid sequences in the hybrid polypeptide comprise different epitopes.

The hybrid polypeptide may comprise one or more polypeptide sequences from the first antigen group and one or more polypeptide sequences from the second antigen group. Accordingly, the invention  
30 includes a composition comprising a first amino acid sequence and a second amino acid sequence, said first amino acid sequence selected from a GAS antigen or a fragment thereof from the first antigen group and said second amino acid sequence selected from a GAS antigen or a fragment thereof from the second antigen group. Preferably, the first and second amino acid sequences in the hybrid polypeptide comprise different epitopes.

Hybrids consisting of amino acid sequences from two, three, four, five, six, seven, eight, nine, or ten GAS antigens are preferred. In particular, hybrids consisting of amino acid sequences from two, three, four, or five GAS antigens are preferred.

Different hybrid polypeptides may be mixed together in a single formulation. Within such combinations, a GAS antigen may be present in more than one hybrid polypeptide and/or as a non-hybrid polypeptide. It is preferred, however, that an antigen is present either as a hybrid or as a non-hybrid, but not as both.

Hybrid polypeptides can be represented by the formula  $\text{NH}_2\text{-A-}\{-\text{X-L-}\}_n\text{-B-COOH}$ , wherein: X is an amino acid sequence of a GAS antigen or a fragment thereof from the first antigen group or the second antigen group; L is an optional linker amino acid sequence; A is an optional N-terminal amino acid sequence; B is an optional C-terminal amino acid sequence; and  $n$  is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15.

If a -X- moiety has a leader peptide sequence in its wild-type form, this may be included or omitted in the hybrid protein. In some embodiments, the leader peptides will be deleted except for that of the -X- moiety located at the N-terminus of the hybrid protein *i.e.* the leader peptide of  $X_1$  will be retained, but the leader peptides of  $X_2 \dots X_n$  will be omitted. This is equivalent to deleting all leader peptides and using the leader peptide of  $X_1$  as moiety -A-.

For each  $n$  instances of  $\{-\text{X-L-}\}$ , linker amino acid sequence -L- may be present or absent. For instance, when  $n=2$  the hybrid may be  $\text{NH}_2\text{-X}_1\text{-L}_1\text{-X}_2\text{-L}_2\text{-COOH}$ ,  $\text{NH}_2\text{-X}_1\text{-X}_2\text{-COOH}$ ,  $\text{NH}_2\text{-X}_1\text{-L}_1\text{-X}_2\text{-COOH}$ ,  $\text{NH}_2\text{-X}_1\text{-X}_2\text{-L}_2\text{-COOH}$ , *etc.* Linker amino acid sequence(s) -L- will typically be short (*e.g.* 20 or fewer amino acids *i.e.* 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples comprise short peptide sequences which facilitate cloning, poly-glycine linkers (*i.e.* comprising  $\text{Gly}_n$ , where  $n = 2, 3, 4, 5, 6, 7, 8, 9, 10$  or more), and histidine tags (*i.e.*  $\text{His}_n$ , where  $n = 3, 4, 5, 6, 7, 8, 9, 10$  or more). Other suitable linker amino acid sequences will be apparent to those skilled in the art. A useful linker is GSGGGG, with the Gly-Ser dipeptide being formed from a *Bam*HI restriction site, thus aiding cloning and manipulation, and the  $(\text{Gly})_4$  tetrapeptide being a typical poly-glycine linker.

-A- is an optional N-terminal amino acid sequence. This will typically be short (*e.g.* 40 or fewer amino acids *i.e.* 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples include leader sequences to direct protein trafficking, or short peptide sequences which facilitate cloning or purification (*e.g.* histidine tags *i.e.*  $\text{His}_n$ , where  $n = 3, 4, 5, 6, 7, 8, 9, 10$  or more). Other suitable N-terminal amino acid sequences will be apparent to those skilled in the art. If  $X_1$  lacks its own N-terminus methionine, -A- is preferably an oligopeptide (*e.g.* with 1, 2, 3, 4, 5, 6, 7 or 8 amino acids) which provides a N-terminus methionine.

-B- is an optional C-terminal amino acid sequence. This will typically be short (*e.g.* 40 or fewer amino acids *i.e.* 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1). Examples include sequences to direct protein trafficking, short peptide sequences which facilitate cloning or purification (*e.g.* comprising histidine



tags *i.e.* His<sub>*n*</sub>, where *n* = 3, 4, 5, 6, 7, 8, 9, 10 or more), or sequences which enhance protein stability. Other suitable C-terminal amino acid sequences will be apparent to those skilled in the art.

Most preferably, *n* is 2 or 3.

5 The invention also provides nucleic acid encoding hybrid polypeptides of the invention. Furthermore, the invention provides nucleic acid which can hybridise to this nucleic acid, preferably under "high stringency" conditions (*e.g.* 65°C in a 0.1xSSC, 0.5% SDS solution).

Polypeptides of the invention can be prepared by various means (*e.g.* recombinant expression, purification from cell culture, chemical synthesis, *etc.*) and in various forms (*e.g.* native, fusions, non-glycosylated, lipidated, *etc.*). They are preferably prepared in substantially pure form (*i.e.* 10 substantially free from other GAS or host cell proteins).

Nucleic acid according to the invention can be prepared in many ways (*e.g.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself, *etc.*) and can take various forms (*e.g.* single stranded, double stranded, vectors, probes, *etc.*). They are preferably prepared in substantially pure form (*i.e.* substantially free from other GAS or host cell nucleic acids).

15 The term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones (*e.g.* phosphorothioates, *etc.*), and also peptide nucleic acids (PNA), *etc.* The invention includes nucleic acid comprising sequences complementary to those described above (*e.g.* for antisense or probing purposes).

20 The invention also provides a process for producing a polypeptide of the invention, comprising the step of culturing a host cell transformed with nucleic acid of the invention under conditions which induce polypeptide expression.

The invention provides a process for producing a polypeptide of the invention, comprising the step of synthesising at least part of the polypeptide by chemical means.

25 The invention provides a process for producing nucleic acid of the invention, comprising the step of amplifying nucleic acid using a primer-based amplification method (*e.g.* PCR).

The invention provides a process for producing nucleic acid of the invention, comprising the step of synthesising at least part of the nucleic acid by chemical means.

### **Strains**

30 Preferred polypeptides of the invention comprise an amino acid sequence found in an M1, M3 or M18 strain of GAS. The genomic sequence of an M1 GAS strain is reported at Ref. 12. The genomic sequence of an M3 GAS strain is reported at Ref. 13. The genomic sequence of an M18 GAS strain is reported at Ref. 14.

Where hybrid polypeptides are used, the individual antigens within the hybrid (*i.e.* individual -X-moieties) may be from one or more strains. Where *n*=2, for instance, X<sub>2</sub> may be from the same strain

as  $X_1$  or from a different strain. Where  $n=3$ , the strains might be (i)  $X_1=X_2=X_3$  (ii)  $X_1=X_2/X_3$  (iii)  $X_1/X_2=X_3$  (iv)  $X_1/X_2/X_3$  or (v)  $X_1=X_2/X_3$ , etc.

#### **Purification and Recombinant Expression**

The GAS antigens of the invention may be isolated from a *Streptococcus pyogenes*, or they may be recombinantly produced, for instance, in a heterologous host. Preferably, the GAS antigens are prepared using a heterologous host. The heterologous host may be prokaryotic (e.g. a bacterium) or eukaryotic. It is preferably *E.coli*, but other suitable hosts include *Bacillus subtilis*, *Vibrio cholerae*, *Salmonella typhi*, *Salmonella typhimurium*, *Neisseria lactamica*, *Neisseria cinerea*, *Mycobacteria* (e.g. *M.tuberculosis*), yeasts, etc.

Recombinant production of polypeptides is facilitated by adding a tag protein to the GAS antigen to be expressed as a fusion protein comprising the tag protein and the GAS antigen. Such tag proteins can facilitate purification, detection and stability of the expressed protein. Tag proteins suitable for use in the invention include a polyarginine tag (Arg-tag), polyhistidine tag (His-tag), FLAG-tag, Strep-tag, c-myc-tag, S-tag, calmodulin-binding peptide, cellulose-binding domain, SBP-tag, chitin-binding domain, glutathione S-transferase-tag (GST), maltose-binding protein, transcription termination anti-terminiation factor (NusA), *E. coli* thioredoxin (TrxA) and protein disulfide isomerase I (DsbA). Preferred tag proteins include His-tag and GST. A full discussion on the use of tag proteins can be found at Ref. 15.

After purification, the tag proteins may optionally be removed from the expressed fusion protein, i.e., by specifically tailored enzymatic treatments known in the art. Commonly used proteases include enterokinase, tobacco etch virus (TEV), thrombin, and factor  $X_a$ .

#### **Immunogenic compositions and medicaments**

Compositions of the invention are preferably immunogenic compositions, and are more preferably vaccine compositions. The pH of the composition is preferably between 6 and 8, preferably about 7.

The pH may be maintained by the use of a buffer. The composition may be sterile and/or pyrogen-free. The composition may be isotonic with respect to humans.

Vaccines according to the invention may either be prophylactic (i.e. to prevent infection) or therapeutic (i.e. to treat infection), but will typically be prophylactic. Accordingly, the invention includes a method for the therapeutic or prophylactic treatment of a *Streptococcus pyogenes* infection in an animal susceptible to streptococcal infection comprising administering to said animal a therapeutic or prophylactic amount of the immunogenic compositions of the invention. Preferably, the immunogenic composition comprises a combination of GAS antigens, said combination consisting of two to thirty-one GAS antigens of the first antigen group. Preferably, the combination of GAS antigens consists of three, four, five, six, seven, eight, nine, or ten GAS antigens selected from the first antigen group. Preferably, the combination of GAS antigens consists of three, four, or five GAS antigens selected from the first antigen group. Preferably, the combination of GAS antigens includes either or both of GAS 40 and GAS 117.



Alternatively, the invention includes an immunogenic composition comprising a combination of GAS antigens, said combination consisting of two to thirty-one GAS antigens of the first antigen group and one, two, three, or four GAS antigens of the second antigen group. Preferably, the combination consists of three, four, five, six, seven, eight, nine, or ten GAS antigens from the first antigen group.

5 Still more preferably, the combination consists of three, four or five GAS antigens from the first antigen group. Preferably, the combination of GAS antigens includes either or both of GAS 40 and GAS 117. Preferably, the combination of GAS antigens includes one or more variants of the M surface protein.

10 The invention also provides a composition of the invention for use as a medicament. The medicament is preferably able to raise an immune response in a mammal (*i.e.* it is an immunogenic composition) and is more preferably a vaccine.

The invention also provides the use of the compositions of the invention in the manufacture of a medicament for raising an immune response in a mammal. The medicament is preferably a vaccine.

15 The invention also provides for a kit comprising a first component comprising a combination of GAS antigens. In one embodiment, the combination of GAS antigens consists of a mixture of two to thirty-one GAS antigens selected from the first antigen group. Preferably, the combination consists of three, four, five, six, seven, eight, nine, or ten GAS antigens from the first antigen group. Preferably, the combination consists of three, four, or five GAS antigens from the first antigen group. Preferably, the combination includes either or both of GAS 117 and GAS 040.

20 In another embodiment, the kit comprises a first component comprising a combination of GAS antigens consisting of a mixture of two to thirty-one GAS antigens of the first antigen group and one, two, three, or four GAS antigens of the second antigen group. Preferably, the combination consists of three, four, five, six, seven, eight, nine, or ten GAS antigens from the first antigen group. Still more preferably, the combination consists of three, four or five GAS antigens from the first antigen group.

25 Preferably, the combination of GAS antigens includes either or both of GAS 40 and GAS 117. Preferably, the combination of GAS antigens includes one or more variants of the M surface protein.

The invention also provides a delivery device pre-filled with the immunogenic compositions of the invention.

30 The invention also provides a method for raising an immune response in a mammal comprising the step of administering an effective amount of a composition of the invention. The immune response is preferably protective and preferably involves antibodies and/or cell-mediated immunity. The method may raise a booster response.

35 The mammal is preferably a human. Where the vaccine is for prophylactic use, the human is preferably a child (*e.g.* a toddler or infant) or a teenager; where the vaccine is for therapeutic use, the human is preferably a teenager or an adult. A vaccine intended for children may also be administered to adults *e.g.* to assess safety, dosage, immunogenicity, *etc.*

These uses and methods are preferably for the prevention and/or treatment of a disease caused by *Streptococcus pyogenes* (e.g. pharyngitis (such as streptococcal sore throat), scarlet fever, impetigo, erysipelas, cellulitis, septicemia, toxic shock syndrome, necrotizing fasciitis (flesh eating disease) and sequelae (such as rheumatic fever and acute glomerulonephritis)). The compositions may also be effective against other streptococcal bacteria.

One way of checking efficacy of therapeutic treatment involves monitoring GAS infection after administration of the composition of the invention. One way of checking efficacy of prophylactic treatment involves monitoring immune responses against the GAS antigens in the compositions of the invention after administration of the composition.

10 Compositions of the invention will generally be administered directly to a patient. Direct delivery may be accomplished by parenteral injection (e.g. subcutaneously, intraperitoneally, intravenously, intramuscularly, or to the interstitial space of a tissue), or by rectal, oral (e.g. tablet, spray), vaginal, topical, transdermal {e.g. see ref. 16} or transcutaneous {e.g. see refs. 17 & 18}, intranasal {e.g. see ref. 19}, ocular, aural, pulmonary or other mucosal administration.

15 The invention may be used to elicit systemic and/or mucosal immunity.

Dosage treatment can be a single dose schedule or a multiple dose schedule. Multiple doses may be used in a primary immunisation schedule and/or in a booster immunisation schedule. In a multiple dose schedule the various doses may be given by the same or different routes e.g. a parenteral prime and mucosal boost, a mucosal prime and parenteral boost, etc.

20 The compositions of the invention may be prepared in various forms. For example, the compositions may be prepared as injectables, either as liquid solutions or suspensions. Solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection can also be prepared (e.g. a lyophilised composition). The composition may be prepared for topical administration e.g. as an ointment, cream or powder. The composition may be prepared for oral administration e.g. as a tablet or capsule, as a spray, or as a syrup (optionally flavoured). The composition may be prepared for pulmonary administration e.g. as an inhaler, using a fine powder or a spray. The composition may be prepared as a suppository or pessary. The composition may be prepared for nasal, aural or ocular administration e.g. as drops. The composition may be in kit form, designed such that a combined composition is reconstituted just prior to administration to a patient. Such kits may comprise one or more antigens in liquid form and one or more lyophilised antigens.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of antigen(s), as well as any other components, as needed. By 'immunologically effective amount', it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, age, the taxonomic group of individual to be treated (e.g. non-human primate, primate, etc.), the capacity of the individual's immune system to synthesise antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's



assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

***Further components of the composition***

The composition of the invention will typically, in addition to the components mentioned above, comprise one or more 'pharmaceutically acceptable carriers', which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolised macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and lipid aggregates (such as oil droplets or liposomes). Such carriers are well known to those of ordinary skill in the art. The vaccines may also contain diluents, such as water, saline, glycerol, *etc.* Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present. A thorough discussion of pharmaceutically acceptable excipients is available in reference 20.

Vaccines of the invention may be administered in conjunction with other immunoregulatory agents. In particular, compositions will usually include an adjuvant.

Preferred further adjuvants include, but are not limited to, one or more of the following set forth below:

**A. Mineral Containing Compositions**

Mineral containing compositions suitable for use as adjuvants in the invention include mineral salts, such as aluminium salts and calcium salts. The invention includes mineral salts such as hydroxides (*e.g.* oxyhydroxides), phosphates (*e.g.* hydroxyphosphates, orthophosphates), sulphates, *etc.* (*e.g.* see chapters 8 & 9 of ref. 21)), or mixtures of different mineral compounds, with the compounds taking any suitable form (*e.g.* gel, crystalline, amorphous, *etc.*), and with adsorption being preferred. The mineral containing compositions may also be formulated as a particle of metal salt. See ref. 22.

**B. Oil-Emulsions**

Oil-emulsion compositions suitable for use as adjuvants in the invention include squalene-water emulsions, such as MF59 (5% Squalene, 0.5% Tween 80, and 0.5% Span 85, formulated into submicron particles using a microfluidizer). See ref. 23.

Complete Freund's adjuvant (CFA) and incomplete Freund's adjuvant (IFA) may also be used as adjuvants in the invention.

**C. Saponin Formulations**

Saponin formulations, may also be used as adjuvants in the invention. Saponins are a heterologous group of sterol glycosides and triterpenoid glycosides that are found in the bark, leaves, stems, roots and even flowers of a wide range of plant species. Saponin from the bark of the *Quillaia saponaria* Molina tree have been widely studied as adjuvants. Saponin can also be commercially obtained from *Smilax ornata* (sarsapilla), *Gypsophilla paniculata* (brides veil), and *Saponaria officianalis* (soap root). Saponin adjuvant formulations include purified formulations, such as QS21, as well as lipid formulations, such as ISCOMs.

Saponin compositions have been purified using High Performance Thin Layer Chromatography (HPLC) and Reversed Phase High Performance Liquid Chromatography (RP-HPLC). Specific purified fractions using these techniques have been identified, including QS7, QS17, QS18, QS21, QH-A, QH-B and QH-C. Preferably, the saponin is QS21. A method of production of QS21 is disclosed in U.S. Patent No. 5,057,540. Saponin formulations may also comprise a sterol, such as cholesterol (see WO 96/33739).

Combinations of saponins and cholesterol can be used to form unique particles called Immunostimulating Complexs (ISCOMs). ISCOMs typically also include a phospholipid such as phosphatidylethanolamine or phosphatidylcholine. Any known saponin can be used in ISCOMs. Preferably, the ISCOM includes one or more of Quil A, QHA and QHC. ISCOMs are further described in EP 0 109 942, WO 96/11711 and WO 96/33739. Optionally, the ISCOMS may be devoid of additional detergent. See ref. 24.

A review of the development of saponin based adjuvants can be found at ref. 25.

#### C. Virosomes and Virus Like Particles (VLPs)

Virosomes and Virus Like Particles (VLPs) can also be used as adjuvants in the invention. These structures generally contain one or more proteins from a virus optionally combined or formulated with a phospholipid. They are generally non-pathogenic, non-replicating and generally do not contain any of the native viral genome. The viral proteins may be recombinantly produced or isolated from whole viruses. These viral proteins suitable for use in virosomes or VLPs include proteins derived from influenza virus (such as HA or NA), Hepatitis B virus (such as core or capsid proteins), Hepatitis E virus, measles virus, Sindbis virus, Rotavirus, Foot-and-Mouth Disease virus, Retrovirus, Norwalk virus, human Papilloma virus, HIV, RNA-phages, QB-phage (such as coat proteins), GA-phage, fr-phage, AP205 phage, and Ty (such as retrotransposon Ty protein p1). VLPs are discussed further in WO 03/024480, WO 03/024481, and Refs. 26, 27, 28 and 29. Virosomes are discussed further in, for example, Ref. 30

#### D. Bacterial or Microbial Derivatives

Adjuvants suitable for use in the invention include bacterial or microbial derivatives such as:

##### (1) *Non-toxic derivatives of enterobacterial lipopolysaccharide (LPS)*

Such derivatives include Monophosphoryl lipid A (MPL) and 3-O-deacylated MPL (3dMPL).

3dMPL is a mixture of 3 De-O-acylated monophosphoryl lipid A with 4, 5 or 6 acylated chains. A preferred "small particle" form of 3 De-O-acylated monophosphoryl lipid A is disclosed in EP 0 689 454. Such "small particles" of 3dMPL are small enough to be sterile filtered through a 0.22 micron membrane (see EP 0 689 454). Other non-toxic LPS derivatives include monophosphoryl lipid A mimics, such as aminoalkyl glucosaminide phosphate derivatives e.g. RC-529. See Ref. 31.

##### (2) *Lipid A Derivatives*

Lipid A derivatives include derivatives of lipid A from *Escherichia coli* such as OM-174. OM-174 is described for example in Ref. 32 and 33.

##### (3) *Immunostimulatory oligonucleotides*



Immunostimulatory oligonucleotides suitable for use as adjuvants in the invention include nucleotide sequences containing a CpG motif (a sequence containing an unmethylated cytosine followed by guanosine and linked by a phosphate bond). Bacterial double stranded RNA or oligonucleotides containing palindromic or poly(dG) sequences have also been shown to be immunostimulatory.

- 5 The CpG's can include nucleotide modifications/analogues such as phosphorothioate modifications and can be double-stranded or single-stranded. Optionally, the guanosine may be replaced with an analogue such as 2'-deoxy-7-deazaguanosine. See ref. 34, WO 02/26757 and WO 99/62923 for examples of possible analogue substitutions. The adjuvant effect of CpG oligonucleotides is further discussed in Refs. 35, 36, WO 98/40100, U.S. Patent No. 6,207,646, U.S. Patent No. 6,239,116, and U.S. Patent  
10 No. 6,429,199.

The CpG sequence may be directed to TLR9, such as the motif GTCGTT or TTCGTT. See ref. 37. The CpG sequence may be specific for inducing a Th1 immune response, such as a CpG-A ODN, or it may be more specific for inducing a B cell response, such as a CpG-B ODN. CpG-A and CpG-B ODNs are discussed in refs. 38, 39 and WO 01/95935. Preferably, the CpG is a CpG-A ODN.

- 15 Preferably, the CpG oligonucleotide is constructed so that the 5' end is accessible for receptor recognition. Optionally, two CpG oligonucleotide sequences may be attached at their 3' ends to form "immunomers". See, for example, refs. 40, 41, 42 and WO 03/035836.

(4) *ADP-ribosylating toxins and detoxified derivatives thereof.*

- 20 Bacterial ADP-ribosylating toxins and detoxified derivatives thereof may be used as adjuvants in the invention. Preferably, the protein is derived from *E. coli* (i.e., *E. coli* heat labile enterotoxin "LT"), cholera ("CT"), or pertussis ("PT"). The use of detoxified ADP-ribosylating toxins as mucosal adjuvants is described in WO 95/17211 and as parenteral adjuvants in WO 98/42375. Preferably, the adjuvant is a detoxified LT mutant such as LT-K63.

E. Human Immunomodulators

- 25 Human immunomodulators suitable for use as adjuvants in the invention include cytokines, such as interleukins (e.g. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (e.g. interferon- $\gamma$ ), macrophage colony stimulating factor, and tumor necrosis factor.

F. Bioadhesives and Mucoadhesives

- 30 Bioadhesives and mucoadhesives may also be used as adjuvants in the invention. Suitable bioadhesives include esterified hyaluronic acid microspheres (Ref. 43) or mucoadhesives such as cross-linked derivatives of poly(acrylic acid), polyvinyl alcohol, polyvinyl pyrrolidone, polysaccharides and carboxymethylcellulose. Chitosan and derivatives thereof may also be used as adjuvants in the invention. E.g., ref. 44.

G. Microparticles

- 35 Microparticles may also be used as adjuvants in the invention. Microparticles (i.e. a particle of ~100nm to ~150 $\mu$ m in diameter, more preferably ~200nm to ~30 $\mu$ m in diameter, and most preferably ~500nm to ~10 $\mu$ m in diameter) formed from materials that are biodegradable and non-toxic (e.g. a poly( $\alpha$ -hydroxy acid), a polyhydroxybutyric acid, a polyorthoester, a polyanhydride, a polycaprolactone, etc.), with poly(lactide-co-glycolide) are preferred, optionally treated to have a

negatively-charged surface (e.g. with SDS) or a positively-charged surface (e.g. with a cationic detergent, such as CTAB).

#### H. Liposomes

Examples of liposome formulations suitable for use as adjuvants are described in U.S. Patent No.

5 6,090,406, U.S. Patent No. 5,916,588, and EP 0 626 169.

#### I. Polyoxyethylene ether and Polyoxyethylene Ester Formulations

Adjuvants suitable for use in the invention include polyoxyethylene ethers and polyoxyethylene esters. Ref. 45. Such formulations further include polyoxyethylene sorbitan ester surfactants in combination with an octoxynol (Ref. 46) as well as polyoxyethylene alkyl ethers or ester surfactants  
10 in combination with at least one additional non-ionic surfactant such as an octoxynol (Ref. 47).

Preferred polyoxyethylene ethers are selected from the following group: polyoxyethylene-9-lauryl ether (laureth 9), polyoxyethylene-9-stearyl ether, polyoxyethylene-8-stearyl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether.

#### J. Polyphosphazene (PCPP)

15 PCPP formulations are described, for example, in Ref. 48 and 49.

#### K. Muramyl peptides

Examples of muramyl peptides suitable for use as adjuvants in the invention include N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), and N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-  
20 hydroxyphosphoryloxy)-ethylamine MTP-PE).

#### L. Imidazoquinolone Compounds.

Examples of imidazoquinolone compounds suitable for use adjuvants in the invention include Imiquamod and its homologues, described further in Ref. 50 and 51.

The invention may also comprise combinations of aspects of one or more of the adjuvants identified  
25 above. For example, the following adjuvant compositions may be used in the invention:

(1) a saponin and an oil-in-water emulsion (ref. 52);

(2) a saponin (e.g., QS21) + a non-toxic LPS derivative (e.g., 3dMPL) (see WO  
94/00153);

(3) a saponin (e.g., QS21) + a non-toxic LPS derivative (e.g., 3dMPL) + a cholesterol;

30 (4) a saponin (e.g. QS21) + 3dMPL + IL-12 (optionally + a sterol) (Ref. 53);

combinations of 3dMPL with, for example, QS21 and/or oil-in-water emulsions (Ref. 54);

(5) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-block polymer L121, and thr-MDP, either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion.

35 (6) Ribi<sup>TM</sup> adjuvant system (RAS), (Ribi Immunochem) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of



monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); and

(7) one or more mineral salts (such as an aluminum salt) + a non-toxic derivative of LPS (such as 3dPML).

5 Aluminium salts and MF59 are preferred adjuvants for parenteral immunisation. Mutant bacterial toxins are preferred mucosal adjuvants.

The composition may include an antibiotic.

#### **Further antigens**

10 The compositions of the invention may further comprise one or more additional non-GAS antigens, including additional bacterial, viral or parasitic antigens.

In one embodiment, the GAS antigen combinations of the invention are combined with one or more additional, non-GAS antigens suitable for use in a paediatric vaccine. For example, the GAS antigen combinations may be combined with one or more antigens derived from a bacteria or virus selected from the group consisting of *N. meningitidis* (including serogroup A, B, C, W135 and/or Y),  
15 *Streptococcus pneumoniae*, *Bordetella pertussis*, *Moraxella catarrhalis*, *Tetanus*, *Diphtheria*, Respiratory Syncytial virus ('RSV'), polio, measles, mumps, rubella, and rotavirus.

In another embodiment, the GAS antigen combinations of the invention are combined with one or more additional, non-GAS antigens suitable for use in a vaccine designed to protect elderly or immunocompromised individuals. For example, the GAS antigen combinations may be combined  
20 with an antigen derived from the group consisting of *Enterococcus faecalis*, *Staphylococcus aureus*, *Staphylococcus epidermis*, *Pseudomonas aeruginosa*, *Legionella pneumophila*, *Listeria monocytogenes*, influenza, and Parainfluenza virus ('PIV').

Where a saccharide or carbohydrate antigen is used, it is preferably conjugated to a carrier protein in order to enhance immunogenicity {e.g. refs. 55 to 64}. Preferred carrier proteins are bacterial toxins  
25 or toxoids, such as diphtheria or tetanus toxoids. The CRM<sub>197</sub> diphtheria toxoid is particularly preferred {65}. Other carrier polypeptides include the *N.meningitidis* outer membrane protein {66}, synthetic peptides {67, 68}, heat shock proteins {69, 70}, pertussis proteins {71, 72}, protein D from *H.influenzae* {73}, cytokines {74}, lymphokines, hormones, growth factors, toxin A or B from *C.difficile* {75}, iron-uptake proteins {76}, etc. Where a mixture comprises capsular saccharides from  
30 both serogroups A and C, it may be preferred that the ratio (w/w) of MenA saccharide:MenC saccharide is greater than 1 (e.g. 2:1, 3:1, 4:1, 5:1, 10:1 or higher). Different saccharides can be conjugated to the same or different type of carrier protein. Any suitable conjugation reaction can be used, with any suitable linker where necessary.

35 Toxic protein antigens may be detoxified where necessary e.g. detoxification of pertussis toxin by chemical and/or genetic means.

Where a diphtheria antigen is included in the composition it is preferred also to include tetanus antigen and pertussis antigens. Similarly, where a tetanus antigen is included it is preferred also to include diphtheria and pertussis antigens. Similarly, where a pertussis antigen is included it is preferred also to include diphtheria and tetanus antigens.

- 5 Antigens in the composition will typically be present at a concentration of at least 1 µg/ml each. In general, the concentration of any given antigen will be sufficient to elicit an immune response against that antigen.

As an alternative to using protein antigens in the composition of the invention, nucleic acid encoding the antigen may be used (e.g. refs. 77 to 85). Protein components of the compositions of the  
10 invention may thus be replaced by nucleic acid (preferably DNA e.g. in the form of a plasmid) that encodes the protein.

### **Definitions**

The term "comprising" means "including" as well as "consisting" e.g. a composition "comprising" X may consist exclusively of X or may include something additional e.g. X + Y.

- 15 The term "about" in relation to a numerical value x means, for example,  $x \pm 10\%$ .

References to a percentage sequence identity between two amino acid sequences means that, when aligned, that percentage of amino acids are the same in comparing the two sequences. This alignment and the percent homology or sequence identity can be determined using software programs known in the art, for example those described in section 7.7.18 of reference 86. A preferred alignment is  
20 determined by the Smith-Waterman homology search algorithm using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 2, BLOSUM matrix of 62. The Smith-Waterman homology search algorithm is disclosed in reference 87.

The following example demonstrates one way of preparing recombinant GAS antigens of the invention and testing their efficacy in a murine model.

25 **EXAMPLE 1: Preparation of recombinant GAS antigens of the invention and Demonstration of Efficacy in Murine Model.**

Recombinant GAS proteins corresponding to two or more of the GAS antigens of the first antigen group are expressed as follows.

30 1. Cloning of GAS antigens for expression in E. coli

The selected GAS antigens were cloned in such a way to obtain two different kinds of recombinant proteins: (1) proteins having an hexa-histidine tag at the carboxy-terminus (Gas-His) and (2) proteins having the hexa-histidine tag at the carboxy-terminus and GST at the amino-terminus (Gst-Gas-His). Type (1) proteins were obtained by cloning in a pET21b+vector  
35 (available from Novagen). The type (2) proteins were obtained by cloning in a pGEX-NNH



vector. This cloning strategy allowed for the GAS genomic DNA to be used to amplify the selected genes by PCR, to perform a single restriction enzyme digestion of the PCR products and to clone then simultaneously into both vectors.

(a) *Construction of pGEX-NNH expression vectors*

5 Two couples of complementary oligodeoxyribonucleotides are synthesised using the DNA synthesiser ABI394 (Perkin Elmer) and reagents from Cruachem (Glasgow, Scotland). Equimolar amounts of the oligo pairs (50 ng each oligo) are annealed in T4 DNA ligase buffer (New England Biolabs) for 10 min in a final volume of 50 µl and then left to cool slowly at room temperature. With the described procedure the following DNA linkers are obtained:

10 **gexNN linker**

NdeI   NheI   XmaI   EcoRI   NcoI   Sall   XhoI   SacI

GATCCCATATGGCTAGCCCGGGGAATTCGTCCATGGAGTGAGTCGACTGACTCGAGTGATCGAGCTC

GGTATACCGATCGGGCCCCTTAAGCAGGTACCTCACTCAGCTGACTGAGCTCACTAGCTCGAG

15 **NotI**

CTGAGCGGCCGCATGAA

GACTCGCCGGCGTACTTTCGA

20 **gexNNH linker**

HindIII   NotI   XhoI   Hexa-Histidine

TCGACAAGCTTGC GGCCGCACTCGAGCATCACCATCACCATCACTGAT

GTTTGAACGCCGGCGTGAGCACGTAGAGGTAGTGGTAGTGACTATCGA

25 The plasmid pGEX-KG [K. L. Guan and J. E. Dixon, *Anal. Biochem.* 192, 262 (1991)] is digested with BamHI and HindIII and 100 ng is ligated overnight at 16 °C to the linker gexNN with a molar ratio of 3:1 linker/plasmid using 200 units of T4 DNA ligase (New England Biolabs). After transformation of the ligation product in *E. coli* DH5, a clone containing the pGEX-NN plasmid, having the correct linker, is selected by means of restriction enzyme analysis and DNA sequencing. The new plasmid pGEX-NN is digested with Sall and HindIII and ligated to the linker gexNNH. After 30 transformation of the ligation product in *E. coli* DH5, a clone containing the pGEX-NNH plasmid, having the correct linker, is selected by means of restriction enzyme analysis and DNA sequencing.

(b) *Chromosomal DNA preparation*

35 GAS SF370 strain is grown in THY medium until OD<sub>600</sub> is 0.6-0.8. Bacteria are then centrifuged, suspended in TES buffer with lysozyme (10mg/ml) and mutanolysine (10U/µl) and incubated 1 hr at 37° C. Following treatment of the bacterial suspension with RNAase, Proteinase K and 10% Sarcosyl/EDTA, protein extraction with saturated phenol and phenol/chloroform is carried out. The resulting supernatant is precipitated with Sodium Acetate/Ethanol and the extracted DNA is pelleted by centrifugation, suspended in Tris buffer and kept at -20° C.

(c) *Oligonucleotide design*

Synthetic oligonucleotide primers are designed on the basis of the coding sequence of each GAS antigen using the sequence of *Streptococcus pyogenes* SF370 M1 strain. Any predicted signal peptide is omitted, by deducing the 5' end amplification primer sequence immediately downstream from the predicted leader sequence. For most GAS antigens, the 5' tail of the primers (see Table 1, below) include only one restriction enzyme recognition site (NdeI, or NheI, or SpeI depending on the gene's own restriction pattern); the 3' primer tails (see Table 1) include a XhoI or a NotI or a HindIII restriction site.

5' tails		3' tails	
NdeI	5' GTGCGTCATATG 3'	XhoI	5' GCGTCTCGAG 3'
NheI	5' GTGCGTGCTAGC 3'	NotI	5' ACTCGCTAGCGGCCGC 3'
SpeI	5' GTGCGTACTAGT 3'	HindIII	5' GCGTAAGCTT 3'

Table 1. Oligonucleotide tails of the primers used to amplify genes encoding selected GAS antigens.

As well as containing the restriction enzyme recognition sequences, the primers include nucleotides which hybridize to the sequence to be amplified. The number of hybridizing nucleotides depends on the melting temperature of the primers which can be determined as described [(Breslauer et al., Proc. Nat. Acad. Sci. 83, 3746-50 (1986 ))]. The average melting temperature of the selected oligos is 50-55 °C for the hybridizing region alone and 65-75 °C for the whole oligos. Oligos can be purchased from MWG-Biotech S.p.A. (Firenze, Italy).

(d) *PCR amplification*

The standard PCR protocol is as follows: 50 ng genomic DNA are used as template in the presence of 0,2 µM each primer, 200 µM each dNTP, 1,5 mM MgCl<sub>2</sub>, 1x PCR buffer minus Mg (Gibco-BRL), and 2 units of Taq DNA polymerase (Platinum Taq, Gibco-BRL) in a final volume of 100 µl. Each sample undergoes a double-step amplification: the first 5 cycles are performed using as the hybridizing temperature of one of the oligos excluding the restriction enzyme tail, followed by 25 cycles performed according to the hybridization temperature of the whole length primers. The standard cycles are as follows:

one cycle:

denaturation : 94 °C, 2 min

5 cycles:

denaturation: 94 °C, 30 seconds, hybridization: 71 °C, 50 seconds, elongation: 72 °C, 1 min or 2 min and 40 sec

25 cycles:

denaturation: 94 °C, 30 seconds

hybridization: 70 °C, 50 seconds

elongation: 72 °C, 1 min or 2 min and 40 sec



72 °C, 7 min  
4 °C

The elongation time is 1 min for GAS antigens encoded by ORFs shorter than 2000 bp, and 2 min and 40 seconds for ORFs longer than 2000 bp. The amplifications are performed using a Gene Amp PCR system 9600 (Perkin Elmer).

To check the amplification results, 4 µl of each PCR product is loaded onto 1-1.5 agarose gel and the size of amplified fragments compared with DNA molecular weight standards (DNA markers III or IX, Roche). The PCR products are loaded on agarose gel and after electrophoresis the right size bands are excised from the gel. The DNA is purified from the agarose using the Gel Extraction Kit (Qiagen) following the instruction of the manufacturer. The final elution volume of the DNA is 50 µl TE (10 mM Tris-HCl, 1 mM EDTA, pH 8). One µl of each purified DNA is loaded onto agarose gel to evaluate the yield.

(e) *Digestion of PCR fragments*

One-two µg of purified PCR products are double digested overnight at 37 °C with the appropriate restriction enzymes (60 units of each enzyme) using the appropriate restriction buffer in 100 µl final volume. The restriction enzymes and the digestion buffers are from New England Biolabs. After purification of the digested DNA (PCR purification Kit, Qiagen) and elution with 30 µl TE, 1 µl is subjected to agarose gel electrophoresis to evaluate the yield in comparison to titrated molecular weight standards (DNA markers III or IX, Roche).

(f) *Digestion of the cloning vectors (pET21b+ and pGEX-NNH)*

10 µg of plasmid is double digested with 100 units of each restriction enzyme in 400 µl reaction volume in the presence of appropriate buffer by overnight incubation at 37 °C. After electrophoresis on a 1% agarose gel, the band corresponding to the digested vector is purified from the gel using the Qiagen Qiaex II Gel Extraction Kit and the DNA was eluted with 50 µl TE. The DNA concentration is evaluated by measuring OD<sub>260</sub> of the sample.

(g) *Cloning of the PCR products*

Seventy five ng of the appropriately digested and purified vectors and the digested and purified fragments corresponding to each selected GAS antigen are ligated in final volumes of 10-20 µl with a molar ratio of 1:1 fragment/vector, using 400 units T4 DNA ligase (New England Biolabs) in the presence of the buffer supplied by the manufacturer. The reactions are incubated overnight at 16 °C. Transformation of *E coli* BL21 (Novagen) and *E coli* BL21-DE3 (Novagen) electrocompetent cells is performed using pGEX-NNH ligations and pET21b+ ligations respectively. The transformation procedure is as follows: 1-2 µl the ligation reaction is mixed with 50 µl of ice cold competent cells, then the cells are poured in a gene pulser 0.1 cm electrode cuvette (Biorad). After pulsing the cells in a MicroPulser electroporator (Biorad) following the manufacturer instructions the cells are suspended in 0.95 ml of SOC medium and incubated for 45 min at 37 °C under shaking. 100 and 900 µl of cell suspensions are plated on separate plates of agar LB 100 µg/ml Ampicillin and the plates are

incubated overnight at 37 °C. The screening of the transformants is done by PCR: randomly chosen transformants are picked and suspended in 30 µl of PCR reaction mix containing the PCR buffer, the 4 dNTPs, 1,5 mM MgCl<sub>2</sub>, Taq polymerase and appropriate forward and reverse oligonucleotide primers that are able to hybridize upstream and downstream from the polylinker of pET21b+ or pGEX-NNH vectors. After 30 cycles of PCR, 5 µl of the resulting products are run on agarose gel electrophoresis in order to select for positive clones from which the expected PCR band is obtained. PCR positive clones are chosen on the basis of the correct size of the PCR product, as evaluated by comparison with appropriate molecular weight markers (DNA markers III or IX, Roche).

## 2. Protein expression

PCR positive colonies are inoculated in 3 ml LB 100 µg/ml Ampicillin and grown at 37 °C overnight. 70 µl of the overnight culture is inoculated in 2 ml LB/Amp and grown at 37 °C until OD<sub>600</sub> of the pET clones reached the 0,4-0,8 value or until OD<sub>600</sub> of the pGEX clones reached the 0,8-1 value. Protein expression is then induced by adding 1 mM IPTG (Isopropil β-D thio-galacto-piranoside) to the mini-cultures. After 3 hours incubation at 37 °C the final OD<sub>600</sub> is checked and the cultures are cooled on ice. After centrifugation of 0.5 ml culture, the cell pellet is suspended in 50 µl of protein Loading Sample Buffer (60 mM TRIS-HCl pH 6.8, 5% w/v SDS, 10% v/v glycerin, 0.1% w/v Bromophenol Blue, 100 mM DTT) and incubated at 100 °C for 5 min. A volume of boiled sample corresponding to 0.1 OD<sub>600</sub> culture is analysed by SDS-PAGE and Coomassie Blue staining to verify the presence of induced protein band.

## 3. Purification of the recombinant proteins

Single colonies are inoculated in 25 ml LB 100 µg/ml Ampicillin and grown at 37 °C overnight. The overnight culture is inoculated in 500 ml LB/Amp and grown under shaking at 25 °C until OD<sub>600</sub> 0.4-0.7. Protein expression is then induced by adding 1 mM IPTG to the cultures. After 3.5 hours incubation at 25 °C the final OD<sub>600</sub> is checked and the cultures are cooled on ice. After centrifugation at 6000 rpm (JA10 rotor, Beckman), the cell pellet is processed for purification or frozen at -20° C.

### (a) *Procedure for the purification of soluble His-tagged proteins from E.coli*

- (1) Transfer the pellets from -20°C to ice bath and reconstitute with 10 ml 50 mM NaHPO<sub>4</sub> buffer, 300 mM NaCl, pH 8,0, pass in 40-50 ml centrifugation tubes and break the cells as per the following outline.
- (2) Break the pellets in the French Press performing three passages with in-line washing.
- (3) Centrifuge at about 30-40000 x g per 15-20 min. If possible use rotor JA 25.50 (21000 rpm, 15 min.) or JA-20 (18000 rpm, 15 min.)
- (4) Equilibrate the Poly-Prep columns with 1 ml Fast Flow Chelating Sepharose resin with 50 mM phosphate buffer, 300 mM NaCl, pH 8,0.
- (5) Store the centrifugation pellet at -20°C, and load the supernatant in the columns.
- (6) Collect the flow through.



- (7) Wash the columns with 10 ml (2 ml + 2 ml + 4 ml) 50 mM phosphate buffer, 300 mM NaCl, pH 8.0.
- (8) Wash again with 10 ml 20 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8.0.
- (9) Elute the proteins bound to the columns with 4.5 ml (1.5 ml + 1.5 ml + 1.5 ml) 250 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8.0 and collect the 3 corresponding fractions of ~1.5 ml each. Add to each tube 15 µl DTT 200 mM (final concentration 2 mM)
- (10) Measure the protein concentration of the first two fractions with the Bradford method, collect a 10 µg aliquot of proteins from each sample and analyse by SDS-PAGE. (N.B.: should the sample be too diluted, load 21 µl + 7 µl loading buffer).
- 10 (11) Store the collected fractions at +4°C while waiting for the results of the SDS-PAGE analysis.
- (12) For immunisation prepare 4-5 aliquots of 100 µg each in 0.5 ml in 40% glycerol. The dilution buffer is the above elution buffer, plus 2 mM DTT. Store the aliquots at -20°C until immunisation.

(b) *Purification of His-tagged proteins from Inclusion bodies*

Purifications are carried out essentially according the following protocol:

- 15 (1) Bacteria are collected from 500 ml cultures by centrifugation. If required store bacterial pellets at -20°C. For extraction, resuspend each bacterial pellet in 10 ml 50 mM TRIS-HCl buffer, pH 8.5 on an ice bath.
- (2) Disrupt the resuspended bacteria with a French Press, performing two passages.
- (3) Centrifuge at 35000 x g for 15 min and collect the pellets. Use a Beckman rotor JA 25.50 (21000 rpm, 15 min.) or JA-20 (18000 rpm, 15 min.).
- 20 (4) Dissolve the centrifugation pellets with 50 mM TRIS-HCl, 1 mM TCEP {Tris(2-carboxyethyl)-phosphine hydrochloride, Pierce} , 6M guanidium chloride, pH 8.5. Stir for ~ 10 min. with a magnetic bar.
- (5) Centrifuge as described above, and collect the supernatant.
- 25 (6) Prepare an adequate number of Poly-Prep (Bio-Rad) columns containing 1 ml of Fast Flow Chelating Sepharose (Pharmacia) saturated with Nickel according to manufacturer recommendations.. Wash the columns twice with 5 ml of H<sub>2</sub>O and equilibrate with 50 mM TRIS-HCl, 1 mM TCEP, 6M guanidinium chloride, pH 8.5.
- (7) Load the supernatants from step 5 onto the columns, and wash with 5 ml of 50 mM TRIS-HCl buffer; 1 mM TCEP, 6M urea, pH 8.5
- 30 (8) Wash the columns with 10 ml of 20 mM imidazole, 50 mM TRIS-HCl , 6M urea, 1 mM TCEP, pH 8.5. Collect and set aside the first 5 ml for possible further controls.
- (9) Elute the proteins bound to the columns with 4.5 ml of a buffer containing 250 mM imidazole, 50 mM TRIS-HCl, 6M urea, 1 mM TCEP, pH 8.5. Add the elution buffer in three 1.5 ml aliquots, and
- 35 collect the corresponding 3 fractions. Add to each fraction 15 µl DTT (final concentration 2 mM).
- (10) Measure eluted protein concentration with the Bradford method, and analyse aliquots of ca 10 µg of protein by SDS-PAGE.

(11) Store proteins at -20°C in 40% (v/v) glycerol, 50 mM TRIS-HCl, 2M urea, 0.5 M arginine, 2 mM DTT, 0.3 mM TCEP, 83.3 mM imidazole, pH 8.5.

(c) *Procedure for the purification of GST-fusion proteins from E.coli*

(1) Transfer the bacterial pellets from -20°C to an ice bath and suspend with 7,5 ml PBS, pH 7,4 to which a mixture of protease inhibitors (COMPLETE™ - Boehringer Mannheim, 1 tablet every 25 ml of buffer) has been added.

(2) Transfer to 40-50 ml centrifugation tubes and sonicate according to the following procedure:

a. Position the probe at about 0,5 cm from the bottom of the tube

b. Block the tube with the clamp

c. Dip the tube in an ice bath

d. Set the sonicator as follows: Timer → Hold, Duty Cycle → 55, Out. Control → 6.

e. perform 5 cycles of 10 impulses at a time lapse of 1 minute (i.e. one cycle = 10 impulses + ~45" hold; b. 10 impulses + ~45" hold; c. 10 impulses + ~45" hold; d. 10 impulses + ~45" hold; e. 10 impulses + ~45" hold).

(3) Centrifuge at about 30-40000 x g for 15-20 min. E.g.: use rotor Beckman JA 25.50 at 21000 rpm, for 15 min.

(4) Store the centrifugation pellets at -20°C, and load the supernatants on the chromatography columns, as follows

(5) Equilibrate the Poly-Prep (Bio-Rad) columns with 0,5 ml (≅1 ml suspension) of Glutathione-Sepharose 4B resin, wash with 2 ml (1 + 1) H<sub>2</sub>O, and then with 10 ml (2 + 4 + 4) PBS, pH 7,4.

(6) Load the supernatants on the columns and discard the flow through.

(7) Wash the columns with 10 ml (2 + 4 + 4) PBS, pH 7,4.

(8) Elute the proteins bound to the columns with 4.5 ml of 50 mM TRIS buffer, 10 mM reduced

glutathione, pH 8,0, adding 1.5 ml + 1.5 ml + 1.5 ml and collecting the respective 3 fractions of ~1.5 ml each.

(9) Measure the protein concentration of the first two fractions with the Bradford method, analyse a 10 µg aliquot of proteins from each sample by SDS-PAGE. (N.B.: if the sample is too diluted load 21 µl (+ 7 µl loading buffer).

(10) Store the collected fractions at +4°C while waiting for the results of the SDS-PAGE analysis.

(11) For each protein destined to the immunisation prepare 4-5 aliquots of 100 µg each in 0.5 ml of 40% glycerol. The dilution buffer is 50 mM TRIS.HCl, 2 mM DTT, pH 8,0. Store the aliquots at -20°C until immunisation.

4. Murine Model of Protection from GAS Infection

(a) *Immunization protocol*

Groups of 10 CD1 female mice aged between 6 and 7 weeks are immunized with two or more GAS antigens of the invention, (20 µg of each recombinant GAS antigen), suspended in 100 µl of suitable solution. Each group receives 3 doses at days 0, 21 and 45. Immunization is performed through intra-peritoneal injection of the protein with an equal volume of Complete Freund's Adjuvant (CFA) for the



first dose and Incomplete Freund's Adjuvant (IFA) for the following two doses. In each immunization scheme negative and positive control groups are used.

For the negative control group, mice are immunized with *E. coli* proteins eluted from the purification columns following processing of total bacterial extract from a *E. coli* strain containing either the pET21b or the pGEX-NNH vector (thus expressing GST only) without any cloned GAS ORF (groups can be indicated as HisStop or GSTStop respectively).

For the positive control groups, mice are immunized with purified GAS M cloned from either GAS SF370 or GAS DSM 2071 strains (groups indicated as 192SF and 192DSM respectively).

Pooled sera from each group is collected before the first immunization and two weeks after the last one. Mice are infected with GAS about a week after.

Immunized mice are infected using a GAS strain different from that used for the cloning of the selected proteins. For example, the GAS strain can be DSM 2071 M23 type, obtainable from the German Collection of Microorganisms and Cell Cultures (DSMZ).

For infection experiments, DSM 2071 is grown at 37° C in THY broth until OD<sub>600</sub> 0.4. Bacteria are pelleted by centrifugation, washed once with PBS, suspended and diluted with PBS to obtain the appropriate concentration of bacteria/ml and administered to mice by intraperitoneal injection. Between 50 and 100 bacteria are given to each mouse, as determined by plating aliquots of the bacterial suspension on 5 THY plates. Animals are observed daily and checked for survival.

#### 5. Analysis of Immune Sera

##### (a) *Preparation of GAS total protein extracts*

Total protein extracts are prepared by incubating a bacterial culture grown to OD<sub>600</sub> 0.4-0.5 in Tris 50mM pH 6.8/mutanolysin (20 units/ml) for 2 hr at 37° C, followed by incubation for ten minutes on ice in 0.24 N NaOH and 0.96% β-mercaptoethanol. The extracted proteins are precipitated by addition of trichloroacetic acid, washed with ice-cold acetone and suspended in protein loading buffer.

##### (b) *Western blot analysis*

Aliquots of total protein extract mixed with SDS loading buffer (1x: 60 mM TRIS-HCl pH 6.8, 5% w/v SDS, 10% v/v glycerin, 0.1% Bromophenol Blue, 100 mM DTT) and boiled 5 minutes at 95° C, were loaded on a 12.5% SDS-PAGE precast gel (Biorad). The gel is run using a SDS-PAGE running buffer containing 250 mM TRIS, 2.5 mM Glycine and 0.1 %SDS. The gel is electroblotted onto nitrocellulose membrane at 200 mA for 60 minutes. The membrane is blocked for 60 minutes with PBS/0.05 % Tween-20 (Sigma), 10% skimmed milk powder and incubated O/N at 4° C with PBS/0.05 % Tween 20, 1% skimmed milk powder, with the appropriate dilution of the sera. After washing twice with PBS/0.05 % Tween, the membrane is incubated for 2 hours with peroxidase-conjugated secondary anti-mouse antibody (Amersham) diluted 1:4000. The nitrocellulose is washed three times for 10 minutes with PBS/0.05 % Tween and once with PBS and thereafter developed by Opti-4CN Substrate Kit (Biorad).

##### (c) *Preparation of Paraformaldehyde treated GAS cultures*

A bacterial culture grown to OD<sub>600</sub> 0.4-0.5 is washed once with PBS and concentrated four times in PBS/0.05 % Paraformaldehyde. Following 1 hr incubation at 37° C with shaking, the treated culture is kept overnight at 4° C and complete inactivation of bacteria is then controlled by plating aliquots on THY blood agar plates.

5 (d) *FACS analysis of Paraformaldehyde treated GAS cultures with mouse immune sera*

About 10<sup>8</sup> Paraformaldehyde inactivated bacteria are washed with 200 µl of PBS in a 96 wells U bottom plate and centrifuged for 10 min. at 3000g, at 4°C. The supernatant is discarded and the bacteria are suspended in 20 µl of PBS-0.1%BSA. Eighty µl of either pre-immune or immune mouse sera diluted in PBS-0.1%BSA are added to the bacterial suspension to a final dilution of either 1:100, 10 1:250 or 1:500, and incubated on ice for 30 min. Bacteria are washed once by adding 100 µl of PBS-0.1%BSA, centrifuged for 10 min. at 3000g, 4°C, suspended in 200 µl of PBS-0.1%BSA, centrifuged again and suspended in 10 µl of Goat Anti-Mouse IgG, F(ab')<sub>2</sub> fragment specific-R-Phycoerythrin-conjugated (Jackson ImmunoResearch Laboratories Inc., cat.N°115-116-072) in PBS-0.1%BSA to a final dilution of 1:100, and incubated on ice for 30 min. in the dark. Bacteria are washed once by 15 adding 180 µl of PBS-0.1%BSA and centrifuged for 10 min. at 3000g, 4°C. The supernatant is discarded and the bacteria were suspended in 200 µl of PBS. Bacterial suspension is passed through a cytometric chamber of a FACS Calibur (Becton Dickinson, Mountain View, CA USA) and 10.000 events are acquired. Data are analysed using Cell Quest Software (Becton Dickinson, Mountain View, CA USA) by drawing a morphological dot plot (using forward and side scatter parameters) on 20 bacterial signals. An histogram plot is then created on FL2 intensity of fluorescence log scale recalling the morphological region of bacteria.

It will be understood that the invention has been described by way of example only and modifications may be made whilst remaining within the scope and spirit of the invention.



## REFERENCES (the contents of which are hereby incorporated by reference)

- 1 Ferretti et al., "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*", PNAS (2001) 98(8):4658 – 4663.
- 2 Beres et al., "Genome sequence of a serotype M3 strain of group A *Streptococcus*: Phage-encoded toxins, the high virulence phenotype, and clone emergence", PNAS (2002) 99(15):10078 – 10083.
- 3 Smoet et al., "Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks", PNAS (2002) 99(7):4668 – 4673.
- 4 Hu et al., "Immunogenicity of a 26-Valent Group A Streptococcal Vaccine" Infection & Immunity (2002) 70(4):2171 – 2177.
- 5 Dale, "Multivalent group A streptococcal vaccine designed to optimize the immunogenicity of six tandem M protein fragments", Vaccine (1999) 17:193 – 200.
- 6 Dale et al., "Recombinant, octavalent group A streptococcal M protein vaccine" Vaccine 14(10):944 – 948.
- 7 Schulze et al., "Stimulation of long-lasting protection against *Streptococcus pyogenes* after intranasal vaccination with non-adjuvanted fibronectin-binding domain of the SfbI protein", Vaccine (2003) 21:1958 – 1964.
- 8 Schulze et al., "Characterization of the Domain of Fibronectin-Binding Protein I of *Streptococcus pyogenes* Responsible for Elicitation of a Protective Immune Response" Infection and Immunity (2001) 69(1):622 – 625.
- 9 Guzman et al., "Protective Immune Response against *Streptococcus pyogenes* in Mice after Intranasal Vaccination with the Fibronectin-binding protein SfbI", Journal of Infectious Diseases (1999) 179:901 – 906.
- 10 Lei et al., "Identification and Characterization of a Novel Heme-Associated Cell Surface Protein Made by *Streptococcus pyogenes*", Infection and Immunity (2002) 70(8):4494 – 4500.
- 11 Dale et al., "Antibodies against a Synthetic Peptide of Saga Neutralize the Cytolytic Activity of Streptolysin S from Group A Streptococci", Infection and Immunity (2002) 70(4):2166 – 2170.
- 12 Ferretti et al., "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*", PNAS (2001) 98(8):4658 – 4663.
- 13 Beres et al., "Genome sequence of a serotype M3 strain of group A *Streptococcus*: Phage-encoded toxins, the high virulence phenotype, and clone emergence", PNAS (2002) 99(15):10078 – 10083.
- 14 Smoet et al., "Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks", PNAS (2002) 99(7):4668 – 4673.
- 15 Terpe et al., "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems", Appl Microbiol Biotechnol (2003) 60:523 – 533.
16. WO99/27961.
17. WO02/074244.
18. WO02/064162.
19. WO03/028760.
20. Gennaro (2000) *Remington: The Science and Practice of Pharmacy*. 20th ed., ISBN: 0683306472.
21. *Vaccine design: the subunit and adjuvant approach* (1995) Powell & Newman. ISBN 0-306-44867-X.
22. WO00/23105.
23. WO90/14837.
24. WO00/07621.
25. Barr, et al., "ISCOMs and other saponin based adjuvants", Advanced Drug Delivery Reviews (1998) 32:247 – 271. See also Sjolander, et al., "Uptake and adjuvant activity of orally delivered saponin and ISCOM vaccines", Advanced Drug Delivery Reviews (1998) 32:321 – 338.

26. Niikura et al., "Chimeric Recombinant Hepatitis E Virus-Like Particles as an Oral Vaccine Vehicle Presenting Foreign Epitopes", *Virology* (2002) 293:273 – 280.
27. Lenz et al., "Papillomavirus-Like Particles Induce Acute Activation of Dendritic Cells", *Journal of Immunology* (2001) 5246 – 5355.
28. Pinto, et al., "Cellular Immune Responses to Human Papillomavirus (HPV)-16 L1 Healthy Volunteers Immunized with Recombinant HPV-16 L1 Virus-Like Particles", *Journal of Infectious Diseases* (2003) 188:327 – 338.
29. Gerber et al., "Human Papillomavirus Virus-Like Particles Are Efficient Oral Immunogens when Coadministered with Escherichia coli Heat-Labile Enterotoxin Mutant R192G or CpG", *Journal of Virology* (2001) 75(10):4752 – 4760.
30. Gluck et al., "New Technology Platforms in the Development of Vaccines for the Future", *Vaccine* (2002) 20:B10 – B16.
31. Johnson et al. (1999) *Bioorg Med Chem Lett* 9:2273-2278.
32. Meraldi et al., "OM-174, a New Adjuvant with a Potential for Human Use, Induces a Protective Response with Administered with the Synthetic C-Terminal Fragment 242-310 from the circumsporozoite protein of Plasmodium berghei", *Vaccine* (2003) 21:2485 – 2491.
33. Pajak, et al., "The Adjuvant OM-174 induces both the migration and maturation of murine dendritic cells in vivo", *Vaccine* (2003) 21:836 – 842.
34. Kandimalla, et al., "Divergent synthetic nucleotide motif recognition pattern: design and development of potent immunomodulatory oligodeoxynucleotide agents with distinct cytokine induction profiles", *Nucleic Acids Research* (2003) 31(9): 2393 – 2400.
35. Krieg, "CpG motifs: the active ingredient in bacterial extracts?", *Nature Medicine* (2003) 9(7): 831 – 835.
36. McCluskie, et al., "Parenteral and mucosal prime-boost immunization strategies in mice with hepatitis B surface antigen and CpG DNA", *FEMS Immunology and Medical Microbiology* (2002) 32:179 – 185.
37. Kandimalla, et al., "Toll-like receptor 9: modulation of recognition and cytokine induction by novel synthetic CpG DNAs", *Biochemical Society Transactions* (2003) 31 (part 3): 654 – 658.
38. Blackwell, et al., "CpG-A-Induced Monocyte IFN-gamma-Inducible Protein-10 Production is Regulated by Plasmacytoid Dendritic Cell Derived IFN-alpha", *J. Immunol.* (2003) 170(8):4061 – 4068.
39. Krieg, "From A to Z on CpG", *TRENDS in Immunology* (2002) 23(2): 64 – 65.
40. Kandimalla, et al., "Secondary structures in CpG oligonucleotides affect immunostimulatory activity", *BBRC* (2003) 306:948 – 953.
41. Kandimalla, et al., "Toll-like receptor 9: modulation of recognition and cytokine induction by novel synthetic CpG DNAs", *Biochemical Society Transactions* (2003) 31(part 3):664 – 658.
42. Bhagat et al., "CpG penta- and hexadeoxynucleotides as potent immunomodulatory agents" *BBRC* (2003) 300:853 – 861.
43. Singh et al. (2001) *J. Cont. Rel.* 70:267-276.
44. WO99/27960.
45. WO99/52549.
46. WO01/21207.
47. WO01/21152.
48. Andrianov et al., "Preparation of hydrogel microspheres by coacervation of aqueous polyphosphazene solutions", *Biomaterials* (1998) 19(1 – 3):109 – 115.
49. Payne et al., "Protein Release from Polyphosphazene Matrices", *Adv. Drug. Delivery Review* (1998) 31(3):185 – 196.
50. Stanley, "Imiquimod and the imidazoquinolones: mechanism of action and therapeutic potential" *Clin Exp Dermatol* (2002) 27(7):571 – 577.
51. Jones, "Resiquimod 3M", *Curr Opin Investig Drugs* (2003) 4(2):214 – 218.
52. WO99/11241.



53. WO98/57659.
54. European patent applications 0835318, 0735898 and 0761231.
55. Ramsay *et al.* (2001) *Lancet* 357(9251):195-196.
56. Lindberg (1999) *Vaccine* 17 Suppl 2:S28-36.
57. Buttery & Moxon (2000) *J R Coll Physicians Lond* 34:163-168.
58. Ahmad & Chapnick (1999) *Infect Dis Clin North Am* 13:113-133, vii.
59. Goldblatt (1998) *J. Med. Microbiol.* 47:563-567.
60. European patent 0 477 508.
61. US Patent No. 5,306,492.
62. International patent application WO98/42721.
63. *Conjugate Vaccines* (eds. Cruse *et al.*) ISBN 3805549326, particularly vol. 10:48-114.
64. Hermanson (1996) *Bioconjugate Techniques* ISBN: 0123423368 or 012342335X.
65. *Research Disclosure*, 453077 (Jan 2002)
66. EP-A-0372501
67. EP-A-0378881
68. EP-A-0427347
69. WO93/17712
70. WO94/03208
71. WO98/58668
72. EP-A-0471177
73. WO00/56360
74. WO91/01146
75. WO00/61761
76. WO01/72337
77. Robinson & Torres (1997) *Seminars in Immunology* 9:271-283.
78. Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648.
79. Scott-Taylor & Dalgleish (2000) *Expert Opin Investig Drugs* 9:471-480.
80. Apostolopoulos & Plebanski (2000) *Curr Opin Mol Ther* 2:441-447.
81. Ilan (1999) *Curr Opin Mol Ther* 1:116-120.
82. Dubensky *et al.* (2000) *Mol Med* 6:723-732.
83. Robinson & Pertmer (2000) *Adv Virus Res* 55:1-74.
84. Donnelly *et al.* (2000) *Am J Respir Crit Care Med* 162(4 Pt 2):S190-193.
85. Davis (1999) *Mt. Sinai J. Med.* 66:84-90.
86. *Current Protocols in Molecular Biology* (F.M. Ausubel *et al.*, eds., 1987) Supplement 30.
87. Smith & Waterman (1981) *Adv. Appl. Math.* 2: 482-489.

# **Document made available under the Patent Cooperation Treaty (PCT)**

**International application number:** PCT/US04/024868

**International filing date:** 30 July 2004 (30.07.2004)

**Document type:** Certified copy of priority document

**Document details:** Country/Office: US  
Number: 60/491,822  
Filing date: 31 July 2003 (31.07.2003)

**Date of receipt at the International Bureau:** 12 January 2005 (12.01.2005)

**Remark:** Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland  
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**